Pursuing forbidden beauty

van Veghel, Maarten

DOI:
10.33612/diss.128123609

Pursuing forbidden beauty

Search for the lepton-flavour violating decays $B^0 \to e^{\pm}\mu^{\mp}$ and $B_s^0 \to e^{\pm}\mu^{\mp}$ and study of electron-reconstruction performance at LHCb

Maarten van Veghel

Cover: Display of the pp collision in LHCb from Run 1 containing the most signal-like $B_s^0 \to e^{\pm}\mu^{\mp}$ candidate. Adapted with AI style generator.

ISBN: 978-94-034-2824-6
First edition.

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). The work is carried out at the National Institute of Subatomic Physics (Nikhef) in Amsterdam, The Netherlands.

rijksuniversiteit
groningen

Pursuing forbidden beauty

Search for the lepton-flavour violating decays $B^0 \to e^{\pm}\mu^{\mp}$ and $B_s^0 \to e^{\pm}\mu^{\mp}$
and study of electron-reconstruction performance at LHCb

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 6 juli 2020 om 9.00 uur

door

Maarten Constantijn van Veghel

geboren op 19 maart 1990
te Tubbergen

Promotor
Prof. dr. A. Pellegrino

Copromotor
Dr. ir. C.J.G. Onderwater

Beoordelingscommissie
Prof. dr. ing. B. van Eijk
Prof. dr. K.H.K.J. Jungmann
Prof. dr. R.G.E. Timmermans

Introduction

To understand the workings of Nature, it is not only necessary to model and describe all we see around us, but also to explore phenomena beyond the reach of our senses in order to measure and deduce Nature's fundamental properties and laws, from which our direct environment emerges. In the 20th century, physicists have developed a very successful theory called the Standard Model, to describe all known fundamental particles and their interactions, aside from gravity [1]. The Standard Model is able to describe Nature down to very small length scales and up to high energies. Despite its success, it is at astronomical length and time scales where major problems occur. As established by many observations, the matter content of the Universe, which should ultimately be built up from fundamental particles, appears to be insufficient. For this reason the enigmatic hypothesis of dark matter has been proposed, *i.e.* unknown massive particles that do not (or very weakly) interact with known matter [2]. Another striking feature of the Universe is that it is vastly dominated by matter, as opposed to antimatter, while at the small length scales described by the Standard Model the difference between the two is very small. Aside from these observational anomalies, there is a more fundamental issue: the Standard Model does not describe gravity, while General Relativity does not describe any quantum nature of gravity, which may be needed to account for gravitational interactions of quantum systems.

To tackle these and other outstanding issues, physicists are exploring into uncharted territories of higher energies and smaller length scales to look for hints of more fundamental laws of Nature, *i.e.* they try to see where the Standard Model breaks. To explore higher energy scales in a laboratory setting, there are two approaches: one can collide particles at high energies to directly search for new phenomena at these scales or one can indirectly probe these or higher scales with precision measurements. The latter is based on the notion of virtual particles. These particles exist only within the time scale of an interaction and can alter processes at energies lower than their mass (although their effect is suppressed by their energy scale relative to the energy scale of the process).

One intermediate approach is the study of b-hadrons. These particles consist of at least a beauty quark, or b-quark, which is one of the heaviest fundamental particles of the Standard Model. Its high mass gives less suppression of the effects of virtual particles of higher masses. In addition, it gives them the opportunity to decay to many combinations of other particles, hence many processes can be studied. Of special interest are rare or forbidden decays of these beauty particles, as their study is a sensitive tool to search for contributions of physics at higher energy scales, as they are tiny on an absolute scale, but can be relatively large with respect to what the Standard Model predicts. Since b-hadrons can be produced by the trillions a year at the high-energy proton collider LHC at CERN in Geneva, rare processes can be studied there to great precision.

In the last decade, hints of deviations from the Standard Model have appeared in a class of b-hadron measurements testing the property of the Standard Model called lepton universality, *i.e.* the property that all three types of charged leptons (electrons, muons and taus) interact in the same way with force mediators [3]. A property of the Standard Model that is closely related to lepton universality is the conservation of lepton-family numbers, also known as lepton-flavour conservation. Hypothetical decays of b-hadrons that violate the conservation of lepton flavour are therefore forbidden. Hence, they provide a good testing ground of the Standard Model and help pin down properties of new physics models that can potentially explain the deviations.

The work presented in this dissertation is part of the search for divergences from lepton universality and lepton-flavour conservation in b-hadron decays. It is performed with data from the LHCb experiment, which makes use of collisions generated by the LHC at CERN. In part I of the thesis, a brief overview of both the experiment and the theoretical context is given, with an emphasis on lepton universality and lepton-flavour violation. In part II, the search for the lepton-flavour violating decays of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ is presented. This analysis led to the most precise measurement to date and resulted in the following peer-reviewed publication:

- LHCb, R. Aaij *et al.*, *Search for the lepton-flavour violating decays* $B^0_{(s)} \to e^\pm\mu^\mp$, JHEP **03** (2018) 078, `arXiv:1710.04111` .

In part III, a study of electron-reconstruction performance at LHCb is presented. The study led to the development of a novel method that in the future can reduce systematic uncertainties of measurements involving electrons at LHCb. This is of crucial importance to lepton universality measurements, especially once the statistical uncertainties will go down with future runs and upgrades of the LHCb experiment. This study led to the following peer-reviewed publication:

- LHCb, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, Journal of Instrumentation **14** (2019) P11023, `arXiv:1909.02957` .

Contents

Part I

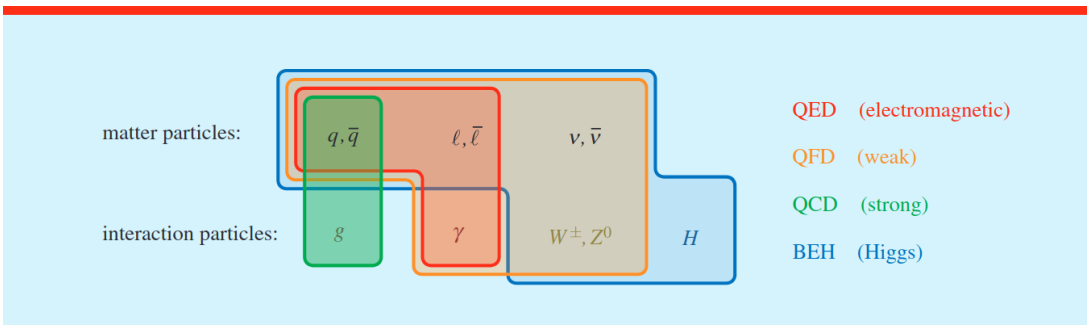
Theoretical context and detector overview

Chapter 1

Physics of lepton-flavour and lepton-universality violation

The current model describing Nature's fundamental particles and their interactions is called the Standard Model [1, 4–8]. The framework of the theory is a quantum field theory [8]. In this context, particles are quantised excitations of fields. The field content consists of spin-$\frac{1}{2}$ particles, called fermions. They are divided into two types: quarks and leptons. The former interact via the strong force and make up hadronic matter like protons, neutrons and also b-hadrons; the latter do not interact with the strong force and consist of two types: charged leptons and neutrinos. As the name suggests, neutrinos do not have charge and subsequently do not interact via the electromagnetic force. All fermions interact via the weak force. Its mediators, the $W^{\pm}$ and Z^0 bosons, are massive. The other force mediators are massless. The masses of the weak force mediators, quarks and charged leptons are generated by spontaneous symmetry breaking of the Higgs field, of which the Higgs boson is an excitation [9–11]. Out of the particles of the Standard Model, the Higgs boson was discovered most recently, in 2012, at the LHC [12, 13]. Note that neutrinos are massless in the Standard Model. The field content of the Standard Model and their interactions are summarised in fig. 1.1.

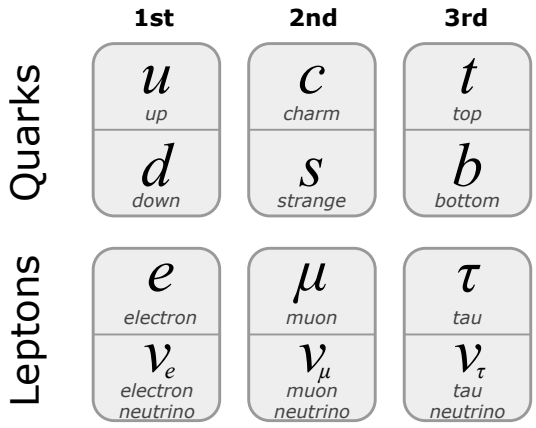
Figure 1.1: Field content of the Standard Model. Quarks, charged leptons and neutrinos are denoted by q, l and ν respectively. Their antimatter counter parts are denoted with a bar. The strong force mediators, gluons, the mediator of the electromagnetic force, the photon, and the Higgs boson are denoted by g, γ and H respectively. Which particle interacts with which forces is demarcated with lines. Illustration from [1].



Quarks and leptons come in three families, where each family consists of two quarks, a charged lepton and a neutrino. The first family, with the lowest masses, make up ordinary

matter like protons, neutrons and electrons. The main difference between the families is the vastly different masses of its particles. The higher the generation, the more massive the quarks and charged leptons are. Since the top quark decays before it can hadronise, the b-hadrons are the heaviest hadrons. Despite of the large mass differences, couplings to the force mediators are the same between families. For leptons, this property is called lepton universality. In addition, the model does not have couplings between different lepton families. Hence lepton flavour, *i.e.* the quantum number associated to leptons of a certain family, is conserved. This is called lepton-flavour conservation. An overview of the different quarks and leptons for all three families is given in fig. 1.2.

Figure 1.2: The three families of quarks and leptons in the Standard Model. The up- and down-quarks make up protons and neutrons. The bottom-quark, *i.e.* beauty quark, makes up b-hadrons together with combinations of the other quarks. The higher the generation, the more massive the quarks and charged leptons are.



Lepton-flavour conservation is tested for charged leptons down to a very high precision, *e.g.* in the search for the lepton-flavour violating decay of $\mu^+ \to e^+\gamma$, which is constrained to have a branching fraction of less than 4×10^{-13} at 90% confidence level [14]. Despite the high precision reached in a number of measurements, to get a complete picture, more measurements are needed, *e.g.* in case hypothetical lepton-flavour violation is not equal among the three families. On the other hand, lepton-flavour conservation for neutrinos is proven to be broken by the discovery of neutrino oscillations [15]. This discovery also implies that neutrinos have mass, which is not possible in the Standard Model, and thus it uncovers the first concrete cracks in the model.

1.1 Flavour anomalies: is lepton universality broken?

Lepton universality is tested to percent level precision for couplings to the weak force mediators or to a higher precision for the electromagnetic interaction, *e.g.* in decays like $J/\psi \to \ell^+\ell^-$ [16]. In contrast, lepton-universality measurements of b-hadron decays in the last decade started to show discrepancies for $b \to s\ell^+\ell^-$ and $b \to c\ell\nu$ transitions, although none of the individual measurement shows the 5σ significance level conventionally required in the particle physics field to claim a new observation [3]. In the Standard

Model, the $b \to s\ell^+\ell^-$ transitions (see *e.g.* fig. 1.3) are heavily suppressed and hence decays like $B^+ \to K^+\ \ell^+\ell^-$, $B^0 \to K^{*0}\ \ell^+\ell^-$ and $\Lambda_b^0 \to pK^-\ \ell^+\ell^-$ are rare, making them an excellent testing ground. Surprisingly, $b \to c\ell\nu$ transitions are far less suppressed but show deviations nonetheless! This hints towards stronger deviations from the Standard Model for higher generations, as $b \to c\ell\nu$ anomalies involve taus versus muons, while $b \to s\ell^+\ell^-$ transition anomalies involve muons versus electrons.

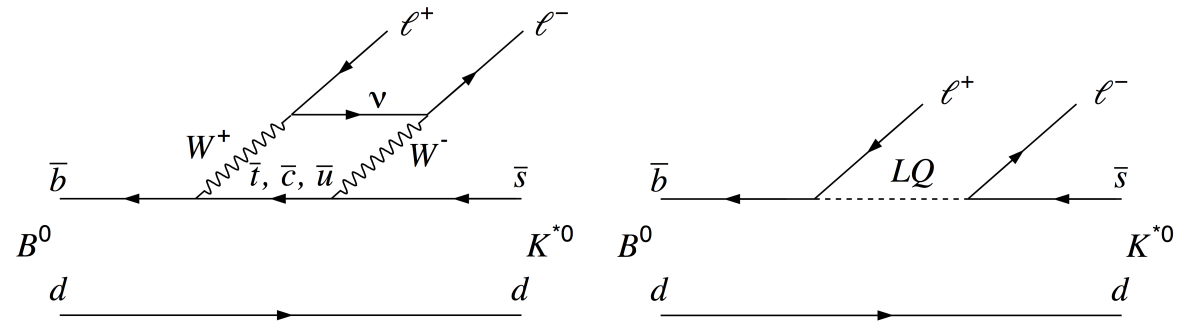
An advantage of studying $b \to s\ell^+\ell^-$ and $b \to c\ell\nu$ transitions in ratios of branching fractions between different leptons is that this is theoretically well predictable, as the strong force is blind towards lepton flavour (aside from the different kinematics due to their masses). Consequently, the uncertainties due to the hadronic part are mostly cancelled in the ratio. Examples of such ratios are,

$$R_X = \frac{\mathcal{B}(X_b \to X\mu^+\mu^-)}{\mathcal{B}(X_b \to Xe^+e^-)}\ , \tag{1.1}$$

where X_b and X can be (combinations of) hadrons like B^+, B^0 or Λ_b^0 and K^+, K^{*0} or pK^- respectively. It is with these ratios that LHCb has measured discrepancies in $b \to s\ell^+\ell^-$ transitions [17–19].

As mentioned before, lepton universality and lepton-flavour violation are closely linked [20]. This can be best illustrated with an example like shown in fig. 1.3. Two Feynman diagrams illustrate first-order contributions in case of the Standard Model and a hypothetical force mediator called a leptoquark [21, 22]. The diagram with the leptoquark can easily give rise to a $b\bar{s} \to e\mu$ transition, which would violate lepton flavour. Given that lepton flavour is conserved in the Standard Model, looking for such forbidden beauty decays would constrain hypothetical explanations of lepton non-universality, while observing one such decay would be an unambiguous sign of new physics.

Figure 1.3: First-order Feynman diagrams of Standard Model (left) and hypothetical leptoquark (right) contributions to the decay of $B^0 \to K^*\ell^+\ell^-$, an example of a $b \to s\ell^+\ell^-$ transition. Illustrations from [18].



Although neutrino oscillations are not part of the Standard Model, it should be noted that these can technically give rise to charged lepton flavour violation as well. This can be shown with the Standard Model diagram in fig. 1.3. If one looks at the internal neutrino line, one can add a neutrino oscillation there, which in turn allows the transition $b \to se\mu$. However, because of the extremely small neutrino masses, the contribution to the amplitude entails a factor of at least 10^{-52} and is therefore way beyond any experimental sensitivity imaginable at present [23, 24].

1.2 The lepton-flavour violating decays of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$

Examples of lepton-flavour violating b-hadron decays are $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$. The most recent search for these decays with the LHCb detector is presented in part II of the thesis.

Hypothetical models that try to explain the flavour anomalies and that also predict a possible enhancement of the branching fractions of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ decays include models with a new Z' gauge boson [25] or leptoquarks [21,26]. Branching fractions can be enhanced up to 10^{-11} in these models. Other models that can also give rise to these decays include heavy singlet Dirac neutrinos [27], supersymmetric models [28] and the Pati-Salam model [29]. A more model-independent theoretical interpretation of $B^0_{(d/s)} \to e^\pm\mu^\mp$ decays is given in [30].

Previous measurements have put limits on the branching fractions of the order of 10^{-8} and 10^{-9} respectively [31]. These branching fractions are defined as,

$$\begin{aligned}\mathcal{B}(B^0_{(d/s)} \to e^\pm\mu^\mp) = \mathcal{B}(B^0_{(d/s)} \to e^+\mu^-) + \mathcal{B}(B^0_{(d/s)} \to e^-\mu^+) \\ + \mathcal{B}(\overline{B}_{(d/s)} \to e^+\mu^-) + \mathcal{B}(\overline{B}_{(d/s)} \to e^-\mu^+)\ ,\end{aligned} \tag{1.2}$$

where each branching fraction is integrated over the decay time of the hadron.

This last notion is important since the decay width of neutral mesons[1] is decay-time dependent due to the possibility of these neutral mesons to transform into their antiparticle through the weak interaction. This is called neutral-meson mixing. Neutral mesons propagate as a superposition of its particle and antiparticle states, or in other words, the interaction eigenstates are different from the eigenstates of the Hamiltonian, *i.e.* the mass eigenstates. Hence, neutral mesons have two mass eigenstates, each with its own mass and lifetime. On the other hand, depending on the properties of interactions, the decay of a neutral particle can occur through the odd, even, or a mixture of the *CP* eigenstates of the two state system, depending on the *CP* properties of the hypothetical mediators. To set model independent limits on the branching fractions, this has to be taken into account while analysing the data [32].

[1] Particles consisting of a quark and an antiquark

Chapter 2

The LHCb detector

To obtain b-hadrons, there are currently two preferred methods. One is to collide electrons with positrons at the center-of-mass energy, $\sqrt{s}$, equivalent to the mass of the $\Upsilon(4S)$ resonance[1] [33, 34]. The vast majority of this resonance decay to b-hadron pairs. Such colliders are therefore called B-factories. The other option is to use a hadron collider. The current most powerful hadron collider is the LHC at CERN in Geneva, Switzerland [35]. While the environment of the B-factories is very clean, as many collisions, when they occur, contain only two b-hadrons, the advantage of a hadron collider is that cross sections, including those for b-hadrons, are much higher. Effectively, hadron colliders produce many more b-hadrons than B-factories. This makes them more suitable for rare decays in case background rejection is sufficient. The latter is the main challenge for detectors operating at hadron machines, as hadron collisions produce many other particles simultaneously with b-hadrons.

The LHCb detector operates at one of the interaction points of the LHC at CERN, designed to study b-hadron and c-hadron decays originating from pp collisions. Many b-hadrons are produced at small polar angles. Hence, the detector is designed as a single-arm forward spectrometer in the pseudo-rapidity range of $2 < \eta < 5$ [36–38]. From a technical perspective, the purpose of the detector is to infer the properties of the particles produced in the collisions: their momentum, origin vertex and type. A schematic overview of the LHCb detector is shown in fig. 2.1.

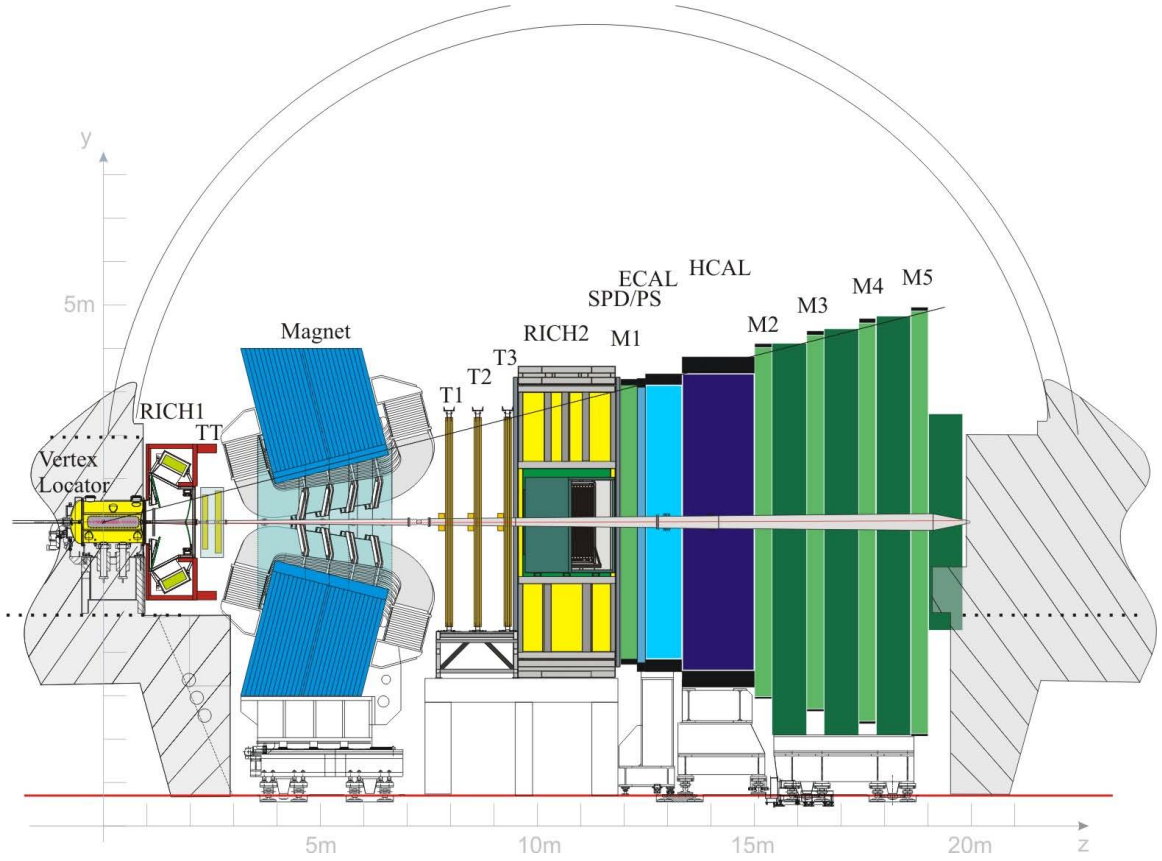
Charged particles[2], like those originating from b-hadron and c-hadron decays travelling in the forward direction, pass through sub-detectors that are designed to accurately track their path while interfering with them the least as possible. One of the goals is to measure their origin with a sub-detector close to the interaction region: the Vertex Locator (VELO). Measuring the origin of tracks is of particular importance for b-hadrons and c-hadrons, since one of their most distinguishing features is their relative long lifetime, as they decay only via the weak interaction. This causes the decay products of these hadrons to be displaced from the primary collision point, called the primary vertex (PV), as depicted in fig. 2.2. Typical distances are of the order of 1 cm. Subsequent decay products have a non-zero impact parameter (IP)[3]. With the high-precision detector modules of the VELO, the IPs of tracks and subsequent displacement of vertices formed by combinations of these tracks can be measured.

[1] Operating at the Z^0 boson mass would also work

[2] *i.e.* charged (pseudo)stable particles: $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, p and $\overline{p}$.

[3] the distance of closest approach of the extrapolated track with respect to the PV.

Figure 2.1: Schematic illustration of the LHCb detector. The opposing beams travel along the z-axis with subsequent collisions happening around $(y, z) = (0, 0)$. The interaction region is surrounded by the Vertex Locator (VELO). The tracking system is completed by the TT and the T-stations (T1,T2,T3) before and after the magnet respectively. Particle identification is done with the RICH1, RICH2, calorimeter system (PS, SPD, ECAL, HCAL) and the muon stations (M1,M2,M3,M4,M5). The x-axis is defined such that it completes a right-handed coordinate system. Note that the bending by the magnet occurs in the xz-plane.



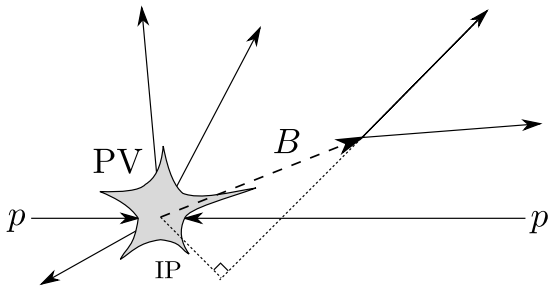
Another goal of the tracking sub-detectors is to obtain the momentum of charged particles. For this purpose, additional tracking sub-detectors have been placed upstream and downstream of the dipole magnet, called the TT and T-stations (T1,T2,T3) respectively. Subsequently, the momenta can be inferred from the deflection and the known magnetic field. The VELO and the additional tracking sub-detectors will be discussed in section 2.1.

Energy deposits of neutral particles, particularly photons, but also of charged particles are detected by the calorimeter system, which is placed after the tracking system. Especially for electrons it is of importance, since due to its low mass it radiates a lot of energy while passing through the detector material. The calorimeter will be discussed in section 2.2.

Muons are particles which do not interact via the strong force and have a mass that is a lot higher than electrons. Hence, they loose the least amount of energy passing through the detector material. Based on this feature, tracking stations, called the muon stations, are placed behind the calorimeter to identify them. The muon stations will be discussed in section 2.3.

To distinguish different charged hadron species, additional detectors are placed between the VELO and the TT and between the T-stations and calorimeter, called the RICH1

Figure 2.2: Schematic illustration of the geometrical features of b-hadron and c-hadron decays. Depicted is a b-hadron decaying to two charged particles. Typical displacements of the decay vertices with respect to the PV are of the order of 1 cm. Charged tracks from these type of decays have an impact parameter (IP) with respect to the PV, which can be obtained by extrapolating the tracks measured with the VELO.



and RICH2 respectively, that measure Cherenkov radiation coming from charged particles travelling through the gas in these detectors. These ring-imaging Cherenkov (RICH) detectors will be discussed in section 2.4.

The LHC at the interaction point of LHCb operated in Run 1 (2011-2012) and Run 2 (2015-2018) with a 40 MHz proton-bunch crossing rate with an instantaneous luminosity of about $2 \cdot 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. During Run 1, the integrated luminosity of pp collisions that was recorded with the LHCb detector corresponds to 1 and $2\,\mathrm{fb}^{-1}$ at $\sqrt{s}$ of 7 and 8 TeV respectively. For Run 2, that was $6\,\mathrm{fb}^{-1}$ at $\sqrt{s}$ of 13 TeV. Under these conditions, the number of produced b-hadrons is of the order of 10^{12} per year[4]. How this large amount of data is processed online, *i.e.* during data taking, and offline will be briefly discussed in sections 2.5 and 2.6 respectively.

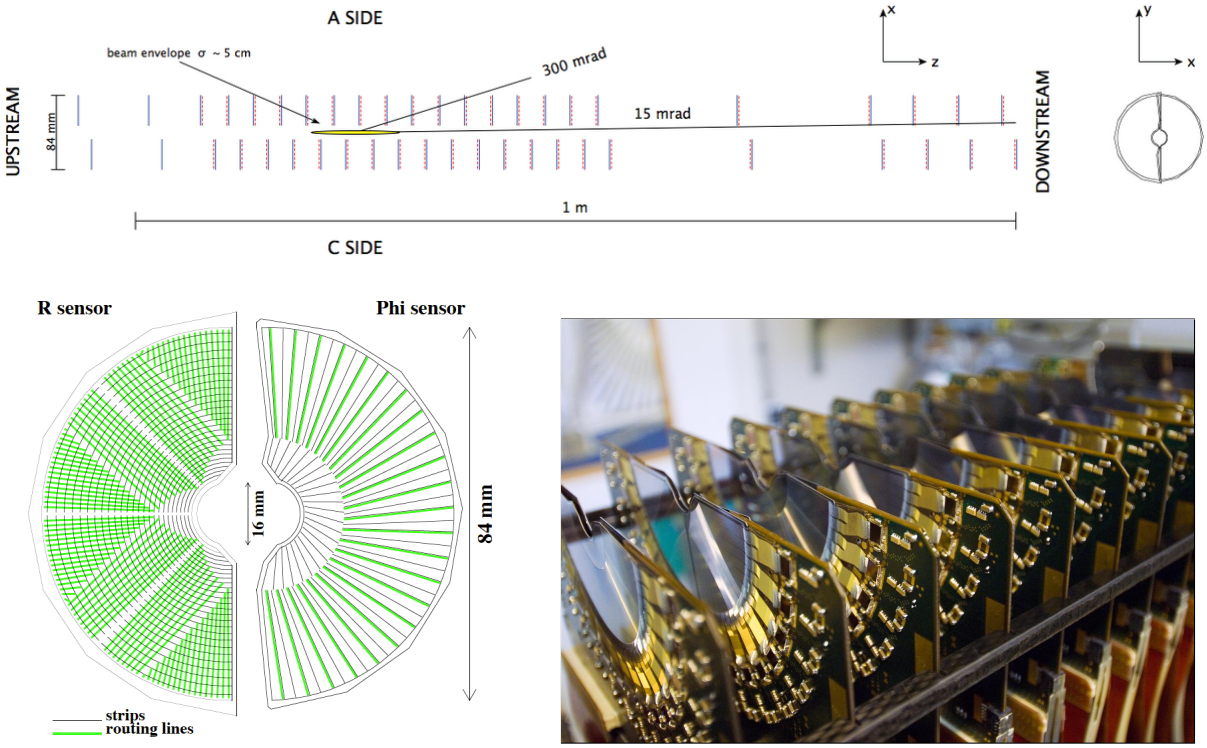
2.1 Tracking and vertex detectors

The detector surrounding the interaction region, the VELO, consists of silicon-strip detector modules of which the surfaces are placed parallel to the beam axis. Its setup and module design is illustrated in fig. 2.3. Charged particles passing through the silicon in the modules alter the conductive properties of the semi-conductor material by ionisation. Passing charged particles kick electrons from the valence band into the conductive band, crossing the small band gap. The electric field applied over the semi-conducting material results in an electric pulse, which allows the hits of these particles to be measured. Modules at specific locations in z consist of a combination of strip sensors measuring r and ϕ coordinates. The modules of the VELO are encased in a thin aluminium foil (the RF foil) to separate the beam vacuum from the VELO vacuum and to act as a Faraday cage to protect the electronics of the VELO from the wake field of the bunches passing by.

The TT consists of large-area silicon detectors placed right before the magnet. It can be used to get a first momentum estimate of the track due to the fringe magnetic field. The main tracking stations (T1,T2,T3) for obtaining the momenta of the tracks are

[4]Note that the b-hadron production cross-sections scale roughly linearly with $\sqrt{s}$ at the energies relevant for the LHC.

Figure 2.3: Overview of the VELO and its modules. The top illustration shows the configuration of the modules. The angles show the range in pseudo-rapidity. Each module consists of an r- and ϕ-coordinate sensor, shown schematically in the bottom left. A photo of the sensors is shown in the bottom right. During data taking, the sensors are about 8 mm close to the beam.



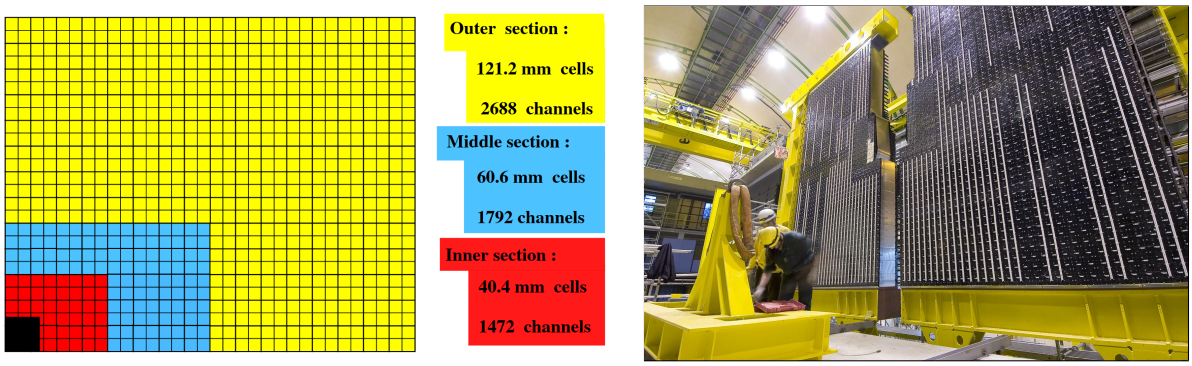
placed after the magnet, making use of the full bending power of the dipole magnet of about 4 Tm. These three stations consist of a combination of two detector technologies. The main one consists of gaseous straw-tube detectors in the outer region, *i.e.* the area away from the beam pipe. This detector is called the Outer Tracker (OT). The position along the tube of hits from passing charged tracks are obtained with the drift time in the electric field of the straw tube of electrons originating from ionisation. A configuration of multiple differently-orientated straw tubes allows to obtain the full set of coordinates of a hit. Around the beam pipe, detectors similar to the TT are placed, collectively called the Inner Tracker (IT). The silicon detectors are better suited to deal with the higher track multiplicity in this region.

2.2 Calorimeter

After the tracking system, both neutral and charged particles pass through the calorimeter system [39]. Its purpose is, in contrast to the tracking system, to interact with its material as much as possible and subsequently deposit energy which can be detected. It starts with the scintillating pad detector (SPD), followed by a layer of lead, the pre-shower detector (PS), the electromagnetic calorimeter (ECAL) and ends with the hadronic calorimeter (HCAL).

The main element is the ECAL. It has a radiation length of 25, which means that for both photons and electrons a very high fraction of its energy will be converted by pair

Figure 2.4: The layout per quadrant and the dimension of the cells in the ECAL are shown on the left. A picture of the calorimeter assembly in the experimental hall is shown on the right.



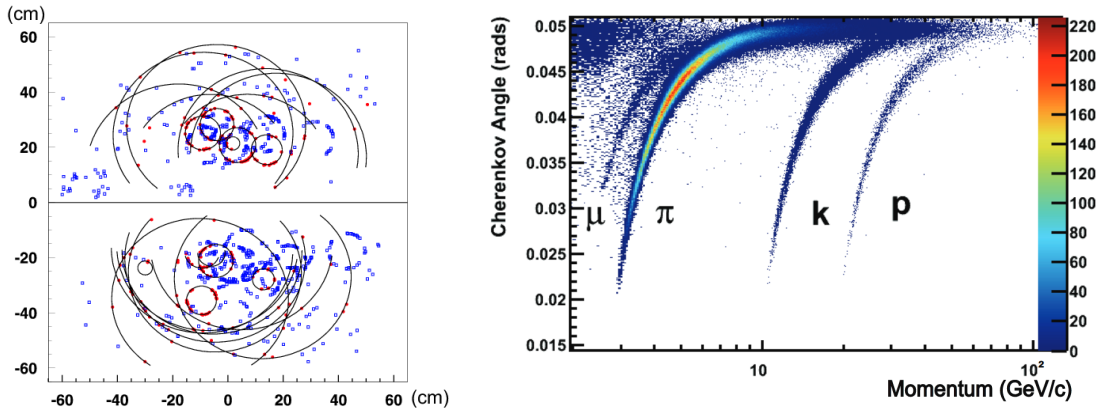
production and bremsstrahlung, causing a shower of electrons and photons. The ECAL consists of cells with alternating layers of lead and scintillating plates. The latter is to detect scintillation light from the calorimeter shower. The cell size and layout is illustrated in fig. 2.4. The SPD adds information if the energy deposit in the ECAL is from a photon or a charged particle, as only charged particles generate scintillation light. The SPD gives a boolean for if the scintillation light was above threshold, called an SPD hit, or not. After the SPD, a layer of lead is placed, giving photons the chance to convert. Early energy deposits from this material, which corresponds to two radiation lengths, are detected by the PS. Like the SPD, the PS consists of scintillating pads. The PS is followed by the ECAL, which in turn is followed by the HCAL, completing the calorimeter system. The HCAL consists of cells of alternating layers of iron and scintillating material. Its purpose is to have additional energy deposits from hadrons, as hadrons tend to not convert all their energy in the ECAL, separating them from electrons and photons. All scintillating materials in the calorimeter are instrumented with photomultipliers. The cell structure throughout the SPD, PS and ECAL are the same. The HCAL has a less granular setup.

Information from the calorimeter is used as input to the identification of different particle species and to reconstruct the momenta of photons by reconstructing clusters of energy deposits in the cells. These clusters can as well be matched to tracks of charged particles, where the latter is reconstructed with hits in the tracking detectors. For example, it is used to recover energy losses of electrons by bremsstrahlung emitted before the magnet, as this affects the momentum determination by the tracking stations that are placed after the magnet. This shows that the reconstruction of energy deposits in the calorimeter is of particular importance to electrons. This will be discussed further in part III. Practically, the particle identification is often used in terms of likelihood ratios between different species for the energy of the deposits in the different parts of the calorimeter.

2.3 Muon stations

Usually, the particles that survive the dense material of the calorimeter are muons. Therefore, the muon stations are placed mainly after the calorimeter to identify them. One station (M1) is placed between the RICH2 and the calorimeter, while the others (M2-M5) are placed after the calorimeter. Iron layers used as absorbers are placed between the stations to further reject hadrons. The detector technology for the tracking stations

Figure 2.5: Rings of photon hits originating from cones of Cherenkov radiation in one of the RICH detectors are shown to the left. The separation of different charged particle species is illustrated by the relation between the cone angle and the momentum shown to the right. Pictures from [41].



are a combination of multiwire proportional chambers in the outer region and triple-GEM detectors in the inner region. Basically, both technologies use the same detection principle as the OT, but in a different configuration, as the muon stations do not need to deal with such a high track multiplicity as the T-stations and subsequently a more cost-effective technology can be used.

Practically, for muon identification, the main discriminant that is used is the boolean `isMuon`. It is obtained by determining if a track from the main tracking system has associable hits in the muon stations or not [40]. Since the penetration power of muons in the muon stations depends on the momentum, a field of interest in the muon stations is determined for a track depending on its momentum per muon station. If there are hits in these fields of interest, `isMuon` is set to true. Like the calorimeter, information from the muon stations is used in the form of likelihood ratios as well.

2.4 RICH

As the calorimeter and muon stations can only distinguish charged particles as electrons, muons or hadrons, more information is needed to identify different hadron species. For this purpose, the two ring-imaging Cherenkov detectors, called RICH1 and RICH2, are placed between the VELO and the TT and the T-stations and the calorimeter respectively. Charged particles passing through a specific gas in these detectors that travel faster than the speed of light in that medium generate Cherenkov radiation, similar to a sonic boom. The cone of light has an angle that is determined by the speed of the charged particle and the refractive index of the gas. With the speed of the particle and the momentum determined by the tracking stations, the mass of the particle can be determined, hence it differentiates between different particle species. Examples of the rings caused by the cones of Cherenkov light and the cone angle versus track momentum are shown in fig. 2.5.

The photons of the Cherenkov radiation are deflected by mirrors to the side of the detector and are subsequently detected by photomultipliers. Rings are not directly reconstructed, but based on the tracks reconstructed by the tracking system, PDFs of the distribution of photons in the detector are constructed based on each particle hypothesis.

With these PDFs and the observed photon hits, likelihood ratios are constructed.

The gas of the RICH1 and RICH2 consists of C_4F_{10} and CF_4 respectively. Hence they cover different momentum ranges and therefore complement each other. Typical ranges are from 1 to 60 GeV/c and from 15 to 100 GeV/c for the RICH1 and RICH2 respectively, which differ slightly per particle type.

2.5 Trigger

Due to a bunch-crossing rate of 40 MHz and a collision rate per bunch-crossing of 1.1, the detector produces large amounts of data. As many collisions contain only backgrounds and since such a large data rate, which would be of the order of 1 TB s^{-1}, is very impractical to save, an online, effectively real-time, selection has to be applied, called a trigger.

The trigger is divided into three stages, where each next stage increases in complexity and decreases in speed. The first stage, called `L0`, is performed by a hardware trigger. Field-programmable gate arrays (FPGAs) are used to make decisions to reduce the rate from 40 MHz to 1 MHz. The main FPGAs of which its passing events take up the majority of the bandwidth are used to do a low-level reconstruction of energy clusters and muon tracks in the calorimeter and muon stations respectively. For the `L0` decisions of the calorimeters, there are three different criteria that can constitute a pass: `L0Photon`, `L0Electron` and `L0Hadron`. All three of the decisions are based on the transverse energy deposit associated to a cluster, given by,

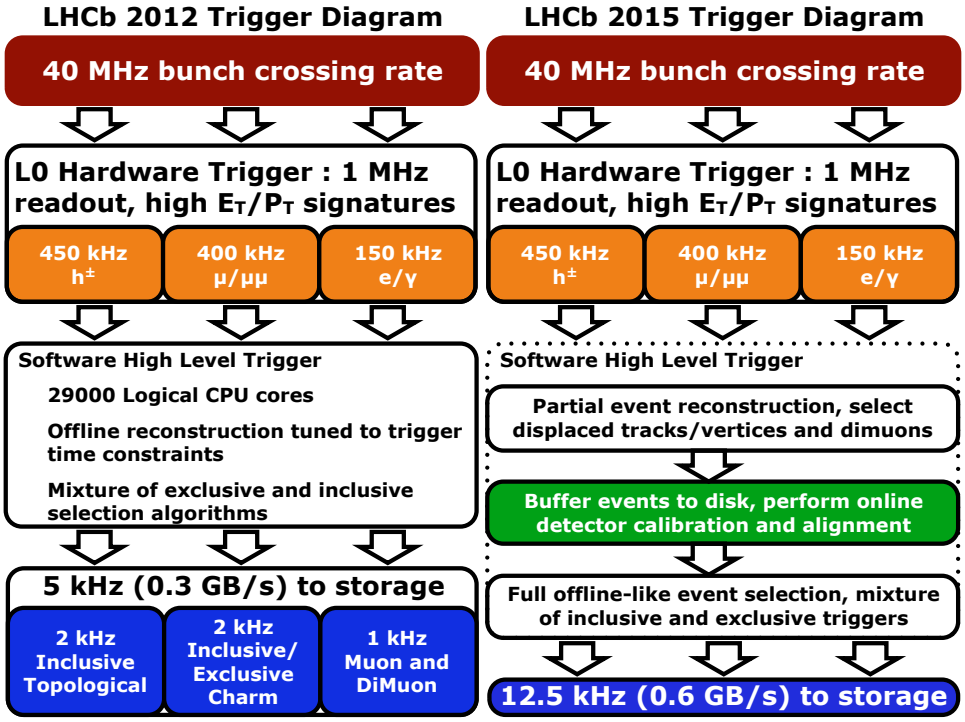
$$E_{\mathrm{T}} = \sum_{i=0}^{4} E_i \sin\theta_i \ , \tag{2.1}$$

where E_i is the energy of the ith cell of the 2×2 cluster and θ_i is the angle between the z-axis and the line from the coordinate origin to the cell. For both `L0Photon` and `L0Electron`, the highest E_{T} cluster in the ECAL must have a certain minimum in E_{T}. The distinction between the two is that `L0Electron` must have an SPD hit in front of the cluster, indicating a charged track in front of the cluster. For `L0Hadron`, the highest E_{T} cluster in the HCAL must have a certain minimum in E_{T}. For the `L0` decisions based on the muon stations, a low-level muon track reconstruction is performed solely with the hits in the muon stations. Straight lines are searched for in each quadrant of the muon stations by the FPGAs. An estimate of the p_{T} of the track is made by assuming it comes from the coordinate origin, taking into account the bending by the magnet. The two highest-p_{T} track candidates per quadrant are used to make a decision if either the track with the highest p_{T} passes a minimum in p_{T} or the combination of the two tracks pass a minimum in $p_{\mathrm{T},1} \times p_{\mathrm{T},2}$, called `L0Muon` and `L0DiMuon` respectively.

For the events that pass the `L0` trigger stage, the data from all sub-detectors are read-out by dedicated electronics and passed to a computer farm. At this point, the software stage starts. Divided in two stages, the first stage reduces the rate from 1 MHz to about 100 kHz. The second stage reduces the rate another order from 100 kHz to about 10 kHz.

The first software stage, called `HLT1`, performs a simplified reconstruction of the tracks in the tracking stations. By starting building tracks in the VELO, which are by good approximation straight lines due the absence of a magnetic field, a selection can already be made on the tracks on their impact parameter and quality. Based on a fairly tight search

Figure 2.6: Trigger levels and their characteristics for Run 1 (left) and Run 2 (right).



window for high momentum tracks, VELO tracks are extrapolated to the TT to obtain a first momentum estimate with the fringe magnetic field. With this information, the search windows in the T-stations can be reduced and subsequent long tracks, *i.e.* tracks with VELO and T-station hits, can be searched for. With the track reconstruction, the PVs in the event can be reconstructed. Aside from track reconstruction, muon particle-identification is performed as well, as it is fast enough to fit in the maximum latency of `HLT1`. Based on the aforementioned reconstruction, for the majority of the available bandwidth in `HLT1`, requirements are set on tracks to be displaced and have a fairly high momentum or to have a higher momentum and no displacement requirement.

The second and last software stage, called `HLT2`, performs a more complete reconstruction, including of the energy deposits in the calorimeter and the particle-identification of the RICH. The bandwidth is allocated to a collection of inclusive and exclusive selections. One of the most important inclusive selections is called the topological trigger. It is designed to efficiently select b-hadron decays with at least two charged particles and will be further discussed in section 3.1.3. Under nominal conditions, the corresponding data rate out of `HLT2` is of the order of $0.1\,\mathrm{GB\,s^{-1}}$. It should be noted that at the `HLT2` level there are quite some differences between Run 1 and Run 2 of the LHC [42]. When necessary, these differences are highlighted in the thesis.

The aforementioned trigger levels and their main characteristics are summarised in fig. 2.6. What should be noted further is that conditions of the LHC change throughout a year of data taking. To optimally make use of this, different settings are used throughout the year.

For the estimation of trigger efficiencies, the association of tracks to calorimeter clusters and muon-station hits is essential. The (mis)association gives rise to a categorisation of

Table 2.1: Description of the categorisation of a trigger decision with respect to the signal. The categories are not mutually exclusive.

`TOS`	The hits left in the detector by the signal are enough to trigger the decision, *i.e.* trigger on signal
`TIS`	The hits originating from the rest of the event are enough to trigger the decision, *i.e.* trigger independent of signal
`TOB`	Only the combination of the hits of the signal and from the rest of the event are enough to trigger the decision, *i.e.* trigger on both

the relation between a trigger decision and signal candidate as given in Table 2.1. Note that these categories are not mutually exclusive.

2.6 Data, reconstruction and simulation

The inference of the original properties of particles from the raw data in the detector, called reconstruction, is already partially discussed in the context of the sub-detectors and the trigger, but there are still some things one should note. Reconstruction can be divided into two categories: finding the signatures of particles in the detector (pattern recognition) and property inference (like the track fit used to obtain the momentum of a charged particle). The performance of these algorithms can be quantified by an efficiency and a resolution for the former and the latter respectively. It is not crucial to know the details of the reconstruction algorithms in general, but a further discussion of reconstruction, with a focus on electrons, is given in chapter 8, to give context to the efficiency measurements presented in part III. Additional algorithms, both applied online and offline, are used to enhance reconstruction performance by improving the alignment and calibration of the sub-detectors.

The performance of the track reconstruction is characterised by three performance numbers, all for minimum ionising particles[5]. First, the reconstruction efficiency for tracks passing the full tracking system is around 96% depending on its kinematics. Second, the momentum resolution ranges from 0.5% to 1.0% from low momentum to about 200 GeV/c respectively. Third, the impact parameter resolution is about $(15 + 29/p_{\mathrm{T}}[\,\mathrm{GeV}/c])\,\mu\mathrm{m}$. In regards of neutral-particle reconstruction, the calorimeter performance can be characterised by its energy resolution of about $10\%/\sqrt{E[\,\mathrm{GeV}]} + 1\%$.

For particle identification, it is useful to build one variable, a test statistic, that has optimal separation between the distribution for signal and background. Subsequently, an analyst can set a requirement on this variable, depending how much background rejection is needed. For the purpose of charged-particle identification, two sets of variables are constructed. One is calculated by combining the individual likelihoods obtained from the calorimeter, muon stations and RICH detectors to form one likelihood ratio with respect to the pion hypothesis, which is the most common charged particle, for each particle species other than the pion. The combination is performed as a sum of the logarithm of the likelihood ratios. These likelihood ratios are called `DLLe`, `DLLmu`, `DLLK` and `DLLp` for the electron, muon, kaon and proton hypotheses respectively. The higher the value of

[5]Practically, all of the (pseudo)stable charged particles, except electrons, fall in this category.

the likelihood, the more signal like it is. At 0, there cannot be made a distinction and the more negative the values, the more background like it is. The other set of variables are constructed with a multivariate classifier. Input variables include likelihood ratios from the individual sub-detectors, kinematic observables and track quality information. The training is performed with simulation, where the true identity of particles is known. The output of the classifiers is transformed into a variable in the range from 0 to 1, where 1 is signal like and 0 is background like. These are called `ProbNNe`, `ProbNNmu`, `ProbNNpi`, `ProbNNk` and `ProbNNp` where the signal is an electron, muon, pion, kaon or proton respectively. Typical performance numbers of the charged-particle identification is of the order of 1% in misidentification probability while having an efficiency of in the 90% for signal.

An important tool for analyses is simulation. At LHCb, including the analyses in this thesis, centrally-produced simulation is used. The simulation makes use of Pythia for the generation of pp collisions [43], with a specific LHCb configuration [44]. The decays of particles orginating from these pp collisions are generated by EvtGen [45], with final-state radiation simulated with Photos [46, 47]. The resulting particles are propagated through a simulated version of the material of the detector to obtain the responses of these particles in all the sub-detectors with Geant4 [48], used as described in [49]. After this, the same tools and algorithms can be used as for data. By keeping track of links between simulated hits in the detector and the originally generated particles, the results of reconstruction, selections and analyses can be traced back to the original particles.

Part II

Search for the lepton-flavour violating decays $B^0 \to e^{\pm}\mu^{\mp}$ and $B_s^0 \to e^{\pm}\mu^{\mp}$

Introduction to Part II

The most recent analysis of the search for the lepton-flavour violating decays of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ with the LHCb detector is presented in this part of the thesis. This work culminated into a peer-reviewed publication [50]. It is based on the dataset collected by the LHCb experiment in 2011 and 2012, corresponding to 1 and $2\,\text{fb}^{-1}$ of pp collisions at $\sqrt{s}$ of 7 and 8 TeV respectively. Hence, it supersedes previous searches of these decays performed by LHCb [31].

Decays are characterised by their branching fraction, *i.e.* the probability for the decaying particle to decay into a specific final state. Therefore, the analysis comes down to measuring the branching fractions or, in case no significant amount is found, to set an upper limit. The subsequent confidence interval can be used by theorists to constrain models. The branching fractions mentioned hereafter are the sum of the four possibilities, *i.e.* the charge-conjugate decays: $B_s^0 \to e^+\mu^-$, $\overline{B}_s^0 \to e^+\mu^-$, $B_s^0 \to e^-\mu^+$ and $\overline{B}_s^0 \to e^-\mu^+$.

Since the B^0 and B_s^0 meson are very similar, with just a small 2% mass difference, the analysis of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ are performed at the same time and with the same setup. As mentioned in section 1.2, a decay of a neutral meson can have an effective lifetime that is specific to that decay due to neutral-meson mixing. This has been taken into account in the analysis of the $B_s^0 \to e^\pm\mu^\mp$ decay. Since the B^0 meson does not have a significant lifetime difference between the heavy and light mass eigenstates, this consideration is not relevant for $B^0 \to e^\pm\mu^\mp$.

To efficiently select the decays of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ in the LHCb dataset while rejecting as much background as possible, a combination of general selections suitable for rare decays and a dedicated multivariate classifier is constructed and is presented in chapter 3. The main consideration in determining these selections is to reject the vast amount of combinatorial background. This background consists of a combination of random electrons and muons that form a good vertex in the detector. While there is a large amount of prompt backgrounds from the hadronic environment of the pp collisions, electrons and muons from separate c-hadron and b-hadron decays, which are also abundant, form a large part of the combinatorial background. The former is mostly rejected by general rare-decay selections, while the latter is the focus of the dedicated multivariate classifier. Aside from the combinatorial backgrounds, misidentified b-hadron decays are rejected by additional requirements.

To convert the signal yield obtained from the candidates passing the selection to branching fractions, a normalisation to well-known b-hadron decays has been performed. In the normalisation ratio systematic uncertainties either cancel or are smaller compared to an absolute measurement. The choice of the normalisation decay channels and the application of the normalisation procedure is discussed in chapter 4.

The yields of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ are extracted from the dataset with a

maximum-likelihood fit to the invariant-mass distribution of the $e^{\pm}\mu^{\mp}$ combination. For signal, the invariant mass should peak at the B^0 or B_s^0 mass, while for background (especially the combinatorial background) it does not. The multivariate classifier is also used to further categorise the dataset in bins of the classifier output. This greatly enhances the sensitivity. Since electrons emit a lot more bremsstrahlung compared to other particles, the dataset is also split into two categories, corresponding to whether bremsstrahlung-photon candidates have been recovered in the detector for the electron candidate or not. This acts as a cross-check for any misidentified b-hadron decays, as the distribution in these categories is vastly different for fake electrons and real electrons. The signal region of the invariant mass was blinded until the selection was fixed, to avoid biases of psychological origin in the choice of the selections. The likelihood fit and the determination of the PDFs for signal and backgrounds is discussed in chapter 5.

To convert the results of the likelihood fit to a confidence interval, a frequentist version of the CLs method is used. The CLs method and its application to this analysis is discussed in chapter 6, together with a brief theoretical interpretation and outlook for future measurements. The study of systematic uncertainties is illustrated throughout the text whenever they are relevant for the discussion.

Chapter 3

Selection

For searches of rare decays in general and thus a fortiori for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$, the main goal of the selection is to put requirements on the dataset such as to keep the efficiency of the signal high and reject as much background as possible. While this might seem obvious, it does not hold for all types of analyses. Rare decays are usually dominated by statistical and not systematic uncertainties, whereas for others it might be the reverse.

The selection starts online, *i.e.* during the data taking when the detector is read out. The event is, at least partially, reconstructed and selections are subsequently performed by the trigger. The trigger selection for $B^0_{(d/s)} \to e^\pm\mu^\mp$ is discussed in 3.1.

The offline selection, which is applied to the events saved by the trigger, has two stages. First, a loose selection is applied based on square cuts on generic variables from reconstruction, discussed in section 3.2. Secondly, a much tighter selection, with highest separation power, is performed using multivariate classifiers and discussed in section 3.3. The main classifier is specifically trained for $B^0_{(d/s)} \to e^\pm\mu^\mp$. Apart from selection, it is also used to categorise the remaining dataset to increase sensitivity, which will be discussed in chapter 5.

As the main focus is on the signal, most of the time only the signal selection is discussed in the coming sections. For datasets used for normalisation or data-driven calibrations, the main goal is to determine a selection that is as close as possible to the signal selection, discussed briefly in appendix B.

3.1 Trigger

During data taking, the selection is done in three stages, as explained in section 2.5. The relevant trigger selections (known as trigger lines) for $B^0_{(d/s)} \to e^\pm\mu^\mp$ and their corresponding thresholds and efficiencies in simulation, used as guidance, will be discussed in this chapter: for `L0` in section 3.1.1, for `HLT1` in section 3.1.2 and for `HLT2` in section 3.1.3.

It should be noted that throughout this chapter, the main goal in designing selections at LHCb, especially at trigger level, is to reject the very large, high multiplicity backgrounds from prompt hadronic interactions and to single out general b-hadron and c-hadron decays and electroweak interactions. Subsequently, these trigger lines are mostly general selections, *i.e.* not specific to $B^0_{(d/s)} \to e^\pm\mu^\mp$. Most threshold are chosen such that the bandwidth is filled, while having as high as possible signal efficiencies for b-hadrons.

Since the majority of the dataset is recorded in 2012, the most common trigger configuration key, hereafter called TCK, `0x00990042`, is used throughout this chapter as

an example. The efficiencies from simulation mentioned in this chapter are all conditional on the loose offline selection (called Stripping), discussed in section 3.2.1.

3.1.1 L0

With a final state of a muon and an electron, the L0 strategy for $B^0_{(d/s)} \to e^\pm\mu^\mp$ has three possibilities: trigger on the signal, *i.e.* TOS, on the muon with L0Muon, on the electron with L0Electron or trigger independent of the signal, *i.e.* TIS, with a combination of L0 triggers.

For L0Muon, the thresholds for the p_T of the muon track, calculated by assuming it originates from the beam spot, are about 1.5 and 1.8 GeV/c for 2011 and 2012 respectively. In case of L0Electron, thresholds are set on the transverse energy deposit in the electromagnetic calorimeter and are around 2.5 and 3.0 GeV for 2011 and 2012 respectively. The thresholds for other calorimeter triggers are for L0Photon the same as and for L0Hadron higher than for L0Electron. For all aforementioned triggers, the number of SPD hits is required to be below 600. This is different for some other lines, and to make a TIS selection, it is best to use a combination of these lines to have uniformity of occupancy in the detector. An example of such selection is TIS on L0Hadron OR L0Muon. The L0 thresholds of relevant lines for common TCK 0x00990042 are listed in table 3.1.

The efficiencies in simulation for $B^0_s \to e^\pm\mu^\mp$ with TCK 0x409f0045, which is close to 0x00990042, of relevant L0 selections are given in Fig. 3.1. It shows that the majority of events, about 75%, are triggered by L0Muon, due to the fairly low effective p_T threshold. The next best, L0Electron, has an efficiency of around 32% and adds relative to L0Muon about 15%. The TIS selection has an efficiency of roughly 28% and adds with respect to L0Muon AND L0Electron just 5%. Since the calibration of such an efficiency is more difficult than for TOS lines, the efficiency is low and the fact that it adds busier events, it is chosen not to include TIS in the selection. Subsequently, the L0 trigger selection for $B^0_{(d/s)} \to e^\pm\mu^\mp$ comprises L0Muon on the muon OR L0Electron on the electron of the $B^0_{(d/s)}$ candidate.

3.1.2 HLT1

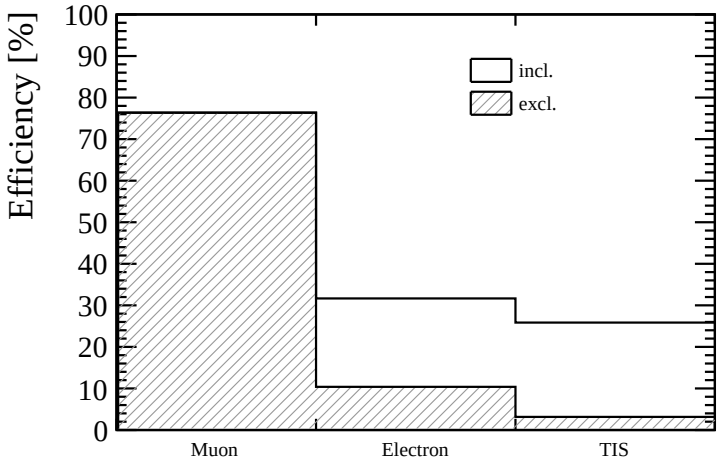
After L0, the detector is read out and subsequently it is possible to perform a simplified but fast version of reconstruction. Reconstruction of VELO tracks is performed and subsequent tracks can be selected based on their impact parameter, IP, with respect to their best PV. Trigger lines using this type of selection are chosen to make up the majority of the HLT1 trigger bandwidth, since they are efficient for decays that are often significantly displaced from the PV and subsequently have a non-zero impact parameter. Decays of b-hadrons and c-hadrons fall into this category, assuming they contain a charged particle in the final state. Therefore, they can be used for $B^0_{(d/s)} \to e^\pm\mu^\mp$ as well.

The main trigger line is Hlt1TrackAllL0. It is based on the available VELO tracks, given that one of the physics lines at L0 has fired. Then, after requiring the IP to be larger than 0.1 mm and tightening track quality requirements, long-track reconstruction is performed and requirements of $p_\mathrm{T} > 1.6\,\mathrm{GeV}/c$ and $p > 3\,\mathrm{GeV}/c$ are set. Next, the track fit is applied to obtain proper uncertainties on the track parameters and with this information, the χ^2/ndf of the track and the difference in χ^2 of including the track in the PV or not, *i.e.* χ^2_IP, is calculated. Requirements are set at track $\chi^2/\mathrm{ndf} < 2$ and

Table 3.1: Requirements of relevant L0 trigger lines for $B^0_{(d/s)} \to e^\pm\mu^\mp$ for L0 TCK 0x0042, which is part of the common 2012 TCK 0x00990042.

Line	Variable	Value	Unit
Muon	n_{SPDHits}	< 600	
	muon track p_{T}	> 1.76	GeV/c
Electron	n_{SPDHits}	< 600	
	E_{T} of electromagnetic calorimeter	> 2.72	GeV
	has SPD hits	True	
	has PS hits	True	
Hadron	n_{SPDHits}	< 600	
	E_{T} of hadronic calorimeter	> 3.62	GeV
Photon	n_{SPDHits}	< 600	
	E_{T} of electromagnetic calorimeter	> 2.72	GeV
	has SPD hits	False	
	has PS hits	True	

Figure 3.1: Efficiencies of a set of relevant L0 trigger lines in simulation of $B^0_s \to e^\pm\mu^\mp$. For L0Muon and L0Electron, TOS is required. For TIS, it is required that at least one of L0Muon or L0Hadron is TIS. For inclusive, the efficiency is just of that specific requirement, while for exclusive, the efficiency is of that specific requirement, given that the other previous requirements, *i.e.* to the left, have not been passed.



$\chi^2_{\text{IP}} > 16$. This trigger line is a good candidate for both the electron and muon candidate of $B^0_{(d/s)} \to e^\pm\mu^\mp$.

Similar to Hlt1TrackAllL0 is Hlt1TrackMuon. The difference is that Hlt1TrackMuon requires, after the long-track reconstruction, that there is a good match with a muon track, *i.e.* isMuon, and after this, the rest of the requirements are set. Due to the extra requirement of isMuon, it is possible to loosen the p_{T} threshold and the track χ^2/ndf to 1 GeV/c and 2.5 respectively. This line is a good candidate for the muon candidate of

Table 3.2: Requirements of relevant HLT1 trigger lines for $B^0_{(d/s)} \to e^\pm\mu^\mp$ for common 2012 TCK 0x00990042. The values in parentheses show the deviation of the TrackMuon line from the TrackAllL0 line.

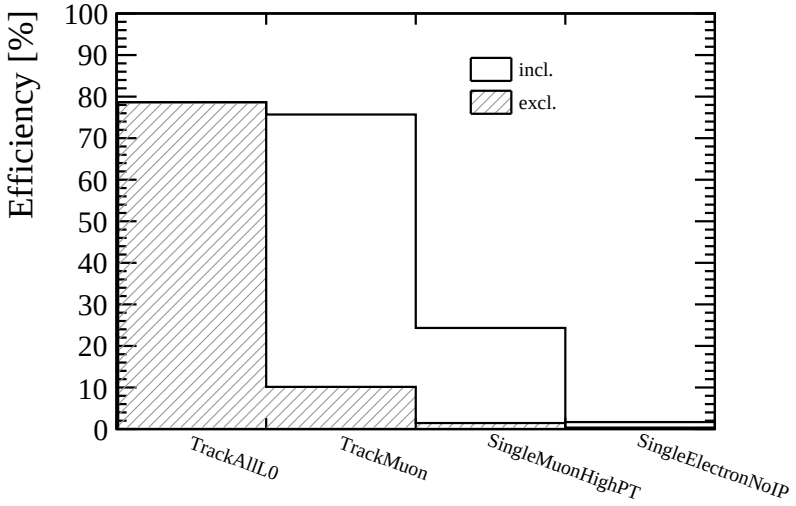
Line	Variable	Value	Unit
TrackAllL0 (TrackMuon)	p_T	$> 1.6\,(1.0)$	GeV/c
	p	> 3.0	GeV/c
	track χ^2/ndf	$< 2\,(2.5)$	
	IP	> 0.1	mm
	χ^2_IP	> 16	
	isMuon	$-$ (True)	
SingleMuonHighPT	p	> 3	GeV/c
	p_T	> 4.8	GeV/c
	track χ^2/ndf	< 3	
	L0*Muon*	DEC	
	isMuon	True	
SingleElectronNoIP	p	> 20	GeV/c
	p_T	> 10	GeV/c
	track χ^2/ndf	< 3	
	L0Electron	DEC	
	E_T of L0Calo match	> 5.08	GeV

$B^0_{(d/s)} \to e^\pm\mu^\mp$.

Apart from lines triggering on displaced tracks, there are single track lines that trigger on prompt tracks, *e.g.* for selecting W-boson decays. Since they don't have the powerful background rejection of requiring displacement, they put extra tight thresholds on p_T and p and a match to L0Electron cluster candidates or isMuon for SingleElectronNoIP and Hlt1SingleMuonHighPT respectively.

The full set of requirements of the Hlt1TrackAllL0, Hlt1TrackMuon, Hlt1SingleMuonHighPT and SingleElectronNoIP lines are given in table 3.2. For $B^0_s \to e^\pm\mu^\mp$, the efficiencies in simulation of using these trigger lines are given in fig. 3.2. Apart from the Stripping selection, they are conditional on the L0 selection defined in section 3.1.1. The line Hlt1TrackAllL0 is the most efficient at around 78%, as it can both trigger on the electron and the muon. Next best, the line Hlt1TrackMuon, is in itself, just triggering on the muon, already at 76%. On top of Hlt1TrackAllL0, it adds about 13%. The prompt lines, Hlt1SingleMuonHighPT and SingleElectronNoIP, would add just 2% due to their strong momentum requirements and are therefore left out. Consequently, the HLT1 trigger selection for $B^0_{(d/s)} \to e^\pm\mu^\mp$, combined with the L0 selection, consists of either L0Muon AND Hlt1TrackMuon on the muon OR L0Electron AND Hlt1TrackAllL0 on the electron of the $B^0_{(d/s)}$ candidate. Cases where the electron triggers L0 but only the muon HLT1 and vice versa adds very little and therefore is left out, to keep the efficiency determination, later discussed in section 4.1.2, simple.

Figure 3.2: Efficiencies of a set of relevant HLT1 trigger lines in simulation of $B_s^0 \to e^{\pm}\mu^{\mp}$. For all lines, TOS is required. In the case of inclusive, the efficiency is just of that specific line, subsequently other lines can be fired as well. For exclusive, the efficiency is of that specific line, given that the other previous lines, *i.e.* to the left, have not fired. Efficiencies are conditional on the L0 selection defined in section 3.1.1.



3.1.3 HLT2: topological trigger

At HLT2 level, the low enough rate allows for a reconstruction that is almost equal to offline reconstruction, *i.e.* in Run 1. Subsequently, higher level requirements can be used to select events, *i.e.* not just based on single-track requirements as in HLT1.

For the purpose of selecting a wide range of b-hadrons, the so-called topological lines have been developed [51, 52]. These lines are based on two-, three- and four-body combinations of displaced charged tracks. The variables used for the selection are chosen to be as inclusive as possible, *i.e.* to trigger on any b-hadron decaying to at least two charged tracks, including neutral particles in the final state and cascade decays where the intermediate particles have a significant decay time. This excludes *e.g.* vertex quality requirements, but a variable such as distance of closest approach, DOCA, is appropriate. To illustrate this, consider the decay $B^+ \to D^0(\to K^-\pi^+)\pi^+$. In this case the final state particles, *i.e.* $K^-\pi^+\pi^+$, do not form a good vertex, because of the lifetime of the D^0, but the two-body combination of $K^-\pi^+$ form a line that should have a small DOCA with respect to the other π^+. Apart from the use of DOCA, using the invariant-mass directly is not efficient if one has *e.g.* neutrals in the final state, but this problem can be alleviated. Due to requiring a large χ^2 associated to the flight distance, FDχ^2, it is possible to know the direction of the b-hadron by the vector pointing from the PV to the decay vertex. Therefore, the missing momentum transverse to the flight direction, p_{miss}, can be determined with the cosine of the angle between the flight direction and the momentum, called DIRA, by $p_{\text{miss}} = p\sqrt{1 - \text{DIRA}^2}$. This missing momentum is added to calculate a corrected invariant-mass, $m_{\text{corr}} = \sqrt{m^2 + p_{\text{miss}}^2} + p_{\text{miss}}$, where m is the invariant mass of the two-, three- or four-body combination.

The input tracks of the topological lines are separated in generic displaced tracks, displaced muons and displaced electrons. Separate lines are constructed where either

Table 3.3: General requirements in the topological triggers lines for common 2012 TCK `0x00990042`. The selection for the electron or muon candidate is on top of the general requirements.

Candidate	Variable	Value	Unit	
General track ($K^{\pm}$)	p_{T}	> 0.5	GeV/c	
	p	> 5	GeV/c	
	χ^2_{IP}	> 4		
	track χ^2/ndf	< 2.5		
$e^{\pm}$	`DLLe`	> -2		
	`L0Electron`	`DEC`		
	`Hlt1Track`* or `Hlt1`*`Electron`	`DEC`		
$\mu^{\pm}$	`isMuon`	`True`		
Combination	DOCA	< 0.2	mm	
	$\sum p_{\mathrm{T}}$	$> 3\,(4)$	GeV/c	2, 3, (4) body
	m	< 7	GeV/c^2	
	FDχ^2	> 100		
	DIRA	> 0		
	`Hlt1Track`*	`TOS`		
One track from combination	p_{T}	$> 1.5\,(1.0)$	GeV/c	$K^{\pm}/e^{\pm}\,(\mu^{\pm})$
	χ^2_{IP}	> 16		

generic charged tracks are used or with at least one muon candidate or an electron candidate. For the muon candidate, `isMuon` is required. For the electron candidate, it is required that `DLLe` > -2 and that `L0Electron` and an `HLT1` track or electron line have fired.

The candidates for the two-, three- and four-body combinations are initially selected as reported in table 3.3, which is quite a loose selection. Subsequently, the tightest requirement is set on a multivariate classifier with DOCA, $\min(p_{\mathrm{T}})$, $\sum p_{\mathrm{T}}$, m, m_{corr} , IPχ^2 as input. The algorithm used to calculate this multivariate classifier is a bonsai boosted decision tree, BBDT [53]. Essentially, it is a boosted decision tree that uses a small number of pre-defined splits the decision tree is allowed to use in order to keep the tree sizes small. Basically, it is using binned input variables. Further explanation of decision trees will be given in section 3.3.3. The use of these small trees reduces computation time, which is essential for running in the online environment of `HLT2`.

The relevant topological lines for $B^0_{(d/s)} \to e^{\pm}\mu^{\mp}$ are `Hlt2Topo2BodyBBDT`, `Hlt2TopoMu2BodyBBDT` and `Hlt2TopoE2BodyBBDT`. As the names suggest, the first one uses generic track candidates, the second one at least one muon candidate and the third one at least one electron candidate as input tracks for the two-body candidates. The requirements specific for these lines are given in table 3.4. Most notable is that the lines with leptons in the final state have a looser selection on the BBDT.

Apart from general topological lines, there are dedicated lines triggering on a b-hadron decaying to two charged final-state particles. A significant difference is that they do not

Table 3.4: Requirements in specific topological triggers lines relevant for $B^0_{(d/s)} \to e^\pm\mu^\mp$ for common 2012 TCK `0x00990042`. Input variables for the `BBDT` are DOCA, $\min(p_\mathrm{T})$, $\sum p_\mathrm{T}$, m, m_corr , IPχ^2. Requirements are the same for 3-body combinations.

	`Topo2BodyBBDT`	`TopoMu2BodyBBDT`	`TopoE2BodyBBDT`
Candidate	$K^\pm K^\pm$ or $K^\pm K^\mp$	$K^\pm\mu^\pm$ or $K^\pm\mu^\mp$	$K^\pm e^\pm$ or $K^\pm e^\mp$
`BBDT`	> 0.4	> 0.1	> 0.1

Table 3.5: Requirements in the `HLT2` trigger line `B2HH` for common 2012 TCK `0x00990042`.

Candidate	Variable	Value	Unit
$\pi^\pm$	p_T	> 1	GeV/c
	IP	> 0.12	mm
	track χ^2	< 3	
B $[\pi^+\pi^-]$	DOCA	< 0.1	mm
	IP	< 0.12	mm
	p_T	> 1.2	GeV/c
	τ_{B^0}	> 0.6	ps
	$m_{\pi^+\pi^-}$	$\in (4.7, 5.9)$	GeV/c^2

use a BBDT, but only one-dimensional requirements. Because of that, they have more stringent momentum requirements. The hypothesis for the invariant-mass calculation is the pion mass. The selection for the `Hlt2B2HH` line is given in table 3.5.

Next to lines for triggering on b-hadron decays, there are lines that select single muons or electrons, as in `HLT1`. They essentially put very stringent requirements on displacement or momentum, or have a pre-scale applied. The selection of relevant single lepton lines are given in table 3.6.

Table 3.6: Requirements in the `HLT2` trigger lines for single muons, `Hlt2SingleMuon` and `Hlt2SingleMuonHighPT`, for common 2012 TCK `0x00990042`.

Line	Variable	Value	Unit
`SingleMuon`	prescale	0.5	
	p_T	> 1.3	GeV/c
	IP	> 0.5	mm
	IPχ^2	> 200	
	track χ^2/ndf	< 2	
	`isMuon`	`True`	
	`Hlt1TrackMuon`	`TOS`	
`SingleMuonHighPT`	prescale	1	
	p_T	> 10	GeV/c
	`isMuon`	`True`	

Figure 3.3: Efficiencies of a set of relevant HLT2 trigger lines in simulation of $B_s^0 \to e^\pm\mu^\mp$. For all lines, TOS is required. For inclusive, the efficiency is just of that specific line, subsequently other lines can be fired as well. For exclusive, the efficiency is of that specific line, given that the other previous lines, *i.e.* to the left, have not fired. Efficiencies are conditional on the L0 and HLT1 selection defined in sections 3.1.1 and 3.1.2 respectively.

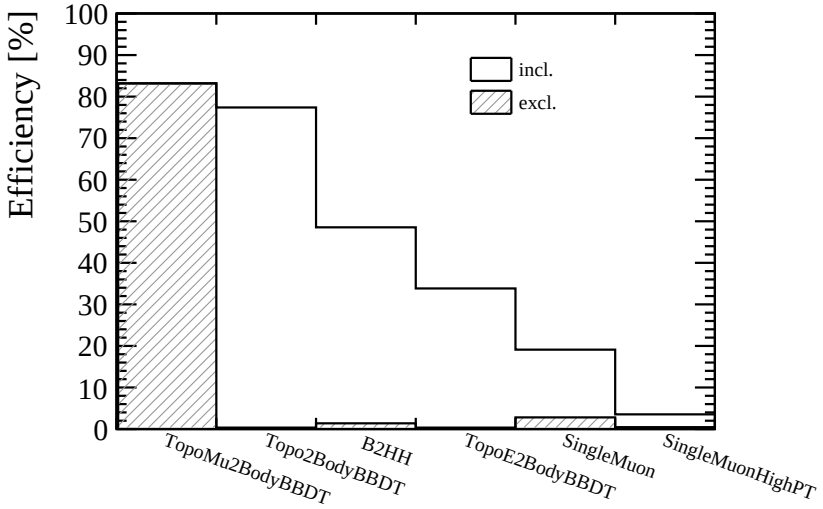


Table 3.7: Requirements in the trigger, at L0, HLT1 and HLT2 level, for $B^0_{(d/s)} \to e^\pm\mu^\mp$. At L0 $\otimes$ HLT1 level, the requirement for the electron, the muon, or both have to be passed.

Candidate	Lines	Value	Note
$e^\pm$	L0Electron and Hlt1TrackAllL0	TOS	or $\mu^\mp$
$\mu^\mp$	L0Muon and	TOS	or $e^\pm$
	Hlt1TrackAllL0 or Hlt1TrackMuon		
$B^0_{(d/s)}$ $[e^\pm\mu^\mp]$	Hlt2TopoMu2BodyBBDT or Hlt2B2HH	TOS	

The efficiencies in simulation of $B_s^0 \to e^\pm\mu^\mp$ of the aforementioned lines are given in fig. 3.3. These are conditional on the L0 and HLT1 selection defined in sections 3.1.1 and 3.1.2 respectively. It shows that the topological lines are most efficient. The most efficient one is the Hlt2TopoMu2BodyBBDT line at around 83%. Adding other topological lines does not add extra efficiency. On top of that, the Hlt2B2HH line adds about 2%. In case of the single muon lines, Hlt2SingleMuon adds about 2% as well. Due to its low additional efficiency and to keep the HLT2 level more similar to control channels, it is chosen to leave those out of the trigger selection for $B^0_{(d/s)} \to e^\pm\mu^\mp$. Therefore, the HLT2 selection for $B^0_{(d/s)} \to e^\pm\mu^\mp$ is given by requiring that at least one of the Hlt2TopoMu2BodyBBDT and Hlt2B2HH lines passes.

3.1.4 L0 $\otimes$ HLT1 $\otimes$ HLT2

The full list of trigger requirements used to select $B_s^0 \to e^\pm\mu^\mp$ signal candidates, as discussed in sections 3.1.1 to 3.1.3, is reported in Tab. 3.7.

3.2 Offline

The offline selection is carried out in two stages. First a loose selection is done, called the Stripping, discussed in 3.2.1. Afterwards some additional fiducial requirements are set, discussed in section 3.2.2. To have enough rejection power of misidentification backgrounds, further particle-identification requirements are set on the electron and muon, discussed in section 3.2.3.

3.2.1 Stripping

Due to the large size of the raw datasets, an extra processing step is needed between the raw data coming out of the trigger and datasets where the analysis is performed on. This step is called the Stripping. The goal is to filter on and save reconstructed objects related to signal channels defined by the various analysis groups, rather than the full raw data. This is done by reconstructing the full event and subsequently combine and select reconstructed objects. Eventually, the objects related to the signal candidates (and only optionally the full events) will be saved in streams of similar lines. These Stripping campaigns are undertaken on collaboration-wide level and take about a month. Afterwards, the analysis groups are able to make datasets within a few days, assuming there is enough grid computing power available.

For rare decays in general, the Stripping output contains only reconstructed objects and variables associated to the signal, and not the full event raw data. This reduces the size per event about an order of magnitude, which in turn allows the selection of the Stripping to be quite loose, as the limiting factor is mostly bandwidth. This reduced data format gives additional freedom for individual analysts to change and optimise their selection on the signal candidates afterwards, albeit only on reconstructed signal objects.

For $B^0_{(d/s)} \to e^{\pm}\mu^{\mp}$, the Stripping selection campaigns `Strip21r1p1` and `Strip21r0p1` are used for 2011 and 2012 respectively. The selection requirements, which are the same for both campaigns, are listed in Table 3.8. For both the electron and muon candidate, it is required that the χ^2_{IP} is larger than 25, essentially a slightly tighter requirement than in the trigger. Further loose requirements are that the p_T of the candidate to be larger than $0.25\,\text{GeV}/c$ and the track χ^2/ndf smaller than 3. Specific for the electron candidate, it is required that the logarithm of the likelihood ratio between electron and pion hypothesis, `DLLe`, is larger than -2. For the muon candidate, the particle identification requires `isMuon` and that the track ghost probability is lower than 0.3. Both particle identification requirements have high efficiencies. These requirements will be superseded by the much tighter ones defined in section 3.2.3.

For the combination of the electron and the muon, the $B^0_{(d/s)}$ candidate, it is required that the vertex has a χ^2 lower than 9 and the DOCA is smaller than 0.3 mm. The candidate should point towards a PV, therefore it is required that the χ^2_{IP} is lower than 25 and that the DIRA, *i.e.* the cosine of the angle between the flight direction and the momentum, is larger than 0. To have enough separation between the PV and the decay vertex candidate, the vertex distance significance, *i.e.* the difference in χ^2 between including the vertex in the PV fit or not, is required to be larger than 225. The invariant mass range of the electron and muon combination is required to be within $1.2\,\text{GeV}/c^2$ of the B^0_s mass.

Table 3.8: Stripping requirements for $B^0_{(d/s)} \to e^\pm\mu^\mp$ of the `Strip21r1p1` and `Strip21r0p1` campaigns used by this analysis.

Candidate	Variable	Value	Unit
$e^\pm$	p_{T}	> 0.25	GeV/c
	χ^2_{IP}	> 25	
	track χ^2/ndf	< 3	
	`InAccPrs` or `InAccEcal`	`True`	
	`DLLe`	> -2	
$\mu^\mp$	p_{T}	> 0.25	GeV/c
	χ^2_{IP}	> 25	
	track χ^2/ndf	< 3	
	track ghost probability	< 0.3	
	`isMuon`	`True`	
$B^0_{(d/s)}$ $[e^\pm\mu^\mp]$	χ^2_{vtx}	< 9	
	DOCA	< 0.3	mm
	χ^2_{IP}	< 25	
	DIRA	> 0	
	χ^2_{VD}	> 225	
	$\lvert m_{e^\pm\mu^\mp} - m_{B^0_s}\rvert$	< 1.2	GeV/c^2

Table 3.9: Pre-selection requirements for $B^0_{(d/s)} \to e^\pm\mu^\mp$.

Candidate	Variable	Value	Unit
$B^0_{(d/s)}$ $[e^\pm\mu^\mp]$	τ	$< 9\,\tau_{B^0_s}$	
	p_{T}	> 0.5	GeV/c
	$m_{e^\pm\mu^\mp}$	$\in (4900, 5850)$	MeV/c^2

3.2.2 Pre-selection

Further offline selection after the Stripping, called pre-selection, is listed in Table 3.9. This consists of applying a maximum to the decay time of 9 times the lifetime of a B^0_s meson and a minimum p_{T} of 0.5 GeV/c for the $B^0_{(d/s)}$ candidate. It also includes the invariant-mass window, $m_{e^\pm\mu^\mp} \in (4900, 5850)\,\mathrm{MeV}/c^2$, used by the likelihood fit to determine the yields of the signal.

3.2.3 Particle identification

The purpose of the majority of the requirements discussed until now is to select a two-body decay from a b-hadron, distinguishing it from mainly combinatorial backgrounds, but also partially reconstructed b-hadron decays. Due to the loose particle-identification requirements in the Stripping and the trigger, the selection barely distinguishes $B^0_{(d/s)} \to e^\pm\mu^\mp$ from decays of a b-hadron to two charged hadrons, denoted by $B \to h^+h^-$, like

Table 3.10: Particle-identification requirements for $B^0_{(d/s)} \to e^\pm\mu^\mp$.

Candidate	Variable	Value	Tuning
$e^\pm$	`DLLe`	> 5.5	
	`ProbNNk`	< 0.95	`MC12TuneV3`
$\mu^\mp$	`ProbNNmu` $(1 -$ `ProbNNk`$)(1 -$ `ProbNNp`$)$	> 0.4	`MC12TuneV2`

$B^0 \to K^+\pi^-$. For this reason, additional particle-identification requirements are set on top of the Stripping selection from table 3.8. The $B \to h^+h^-$ backgrounds will be discussed further in section 5.2.

In case of the muon candidate, the same requirement is used as the $B^0_{(d/s)} \to \mu^+\mu^-$ analysis, since the backgrounds and the kinematics of the muon are very similar [54]. For the electron candidate, to determine the best set of particle identification criteria, a figure of merit of how well the criterium rejects $B \to h^+h^-$ backgrounds is constructed. It is defined to take into account their branching fraction, relative b-meson production fractions and misidentification probabilities, by,

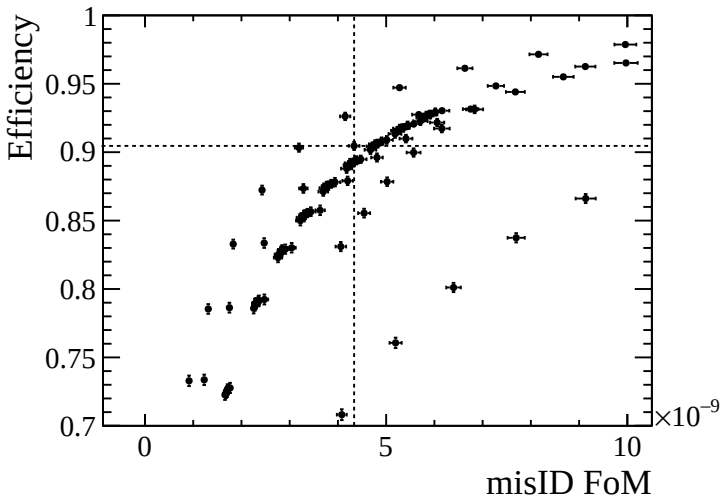
$$\text{misID FoM} = \sum_{B^0_{(d/s)} \to \text{hh}} \frac{f_{d,s}}{f_d} \mathcal{B}(B \to h^+h^-) \epsilon^{\text{PID}}_{\text{hh}\to e\mu} \ , \tag{3.1}$$

where $\frac{f_{d,s}}{f_d}$ is the relative hadron production fraction, $\mathcal{B}(B \to h^+h^-)$ the branching fraction of each $B \to h^+h^-$ decay and $\epsilon^{\text{PID}}_{\text{hh}\to e\mu}$ is the misidentification probability of the specific hadrons to be identified as electrons and muons. The sum is performed over decays of $B^0_{(d/s)}$ to $K^\pm\pi^\mp$, K^+K^- and $\pi^+\pi^-$. To good approximation, *i.e.* within a few percent, the efficiency of the selection without particle identification is the same for each $B \to h^+h^-$ decay. The misidentification probabilities and the efficiency for signal are all calculated according to the data-driven method that will be described in section 4.1.3.

The $B^0_{(d/s)} \to e^\pm\mu^\mp$ particle-identification efficiency is tuned to be around the 2011 analysis efficiency, *i.e.* around 90%. This way, the misidentification rate will go down while keeping the same efficiency, to keep up with the increased sensitivity. The optimisation, by setting the rejection figure-of-merit against the signal efficiency, is depicted in fig. 3.4. Combinations of `DLLe` and `ProbNNe` with versions `MC12TuneV2` and `MC12TuneV3`; `DLLK` and `ProbNNk`, also with versions `MC12TuneV2` and `MC12TuneV3`, are used. The `DLLK` and `ProbNNk` are used to reject kaons and `DLLe` and `ProbNNe` to reject pions. The resulting electron particle-identification criterium is `DLLe` above 5.5 and `ProbNNk`, using the `MC12TuneV3`, below 0.95.

The full set of particle identification requirements for $B^0_{(d/s)} \to e^\pm\mu^\mp$ is given in table 3.10.

Figure 3.4: Optimisation of particle-identification requirement for $B^0_{(d/s)} \to e^\pm\mu^\mp$ with respect to $B \to h^+h^-$ misidentification. The dashed lines represent the particle-identification requirement used in the 2011 analysis. All the efficiencies are calculated with data-driven methods using reweighing to kinematics of $B^0_s \to e^\pm\mu^\mp$ after the Stripping selection.



3.3 Multivariate classification

While the selection up to this point focuses on the suppression of prompt and misidentified backgrounds, there is still much of the largest and toughest background remaining: combinatorial. Real electrons and muons that are displaced from their primary vertex, originating from separate b-hadron and c-hadron decays, can still pass the selection. As many c-hadron and b-hadron decays are produced in pp collisions, many of these backgrounds still remain in the dataset.

To fight these backgrounds, a multivariate classification is used to combine multiple observables into one observable, *i.e.* a test statistic, that has optimised separation power between signal and combinatorial. How this is done is briefly discussed in section 3.3.1.

This separation of signal from background is done in two ways. First, a generic b-hadron decay classifier is trained, called BDTS, that is also applicable to calibration and normalisation channels. The goal is to have a high efficiency for signal while already rejecting a lot of background. This is discussed in section 3.3.2. Secondly, a specific classifier is trained for $B^0 \to e^\pm\mu^\mp$ and $B^0_s \to e^\pm\mu^\mp$, where a stricter selection is applied and the remaining data is categorised in bins of the classifier output. The construction of this classifier is discussed in section 3.3.3. The categorization in bins of the classifier output is discussed later in section 5.1.2.

3.3.1 Test-statistic building

The goal of a multivariate classifier is to build a test statistic that has distributions of signal and background that are maximally separated. The best option is to construct a multidimensional likelihood ratio of signal over background [55]. Often, especially with large numbers of variables, this is difficult to construct and other algorithms are used to approach its optimal performance.

For this purpose, a Boosted Decision Tree (BDT) is used to construct a classifier [56]. At the core lies a decision tree, which is a set of consecutive splits in the dataset in one of the input variables. These decision trees start with a split in the variable that has the largest separation value, the root node. Subsequently, the next sub-node is split in the then best separating variable at the best separating value. It may make a split in the same variable. The splitting is stopped after a maximum amount of splits or when purity is too low. The last split is called a leaf. A set of such splits is called a tree. The leaves are categorised as either signal or background depending on which is the majority. Multiple trees are built according to the dataset where misclassified entries have a higher weight. This is called boosting and it stabilises the response of the final classifier, reducing the chance that decision trees are guided by statistical fluctuations. The latter is called over-training. In the end, a single classifier is built by a weighted average of all the trained trees.

3.3.2 BDTS

The first line of defence against combinatorial backgrounds is BDTS. It uses a BDT. It has been originally developed for another analysis, using the same training [54, 57]. The purpose is to reject a first large set of combinatorial while have a high, about 95%, efficiency for signal. The goal is to also use the BDTS for the normalisation channels. As will be discussed in chapter 4, one of these channels is $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+$. In that case, some of the variables are computed with the two tracks of the J/ψ, to keep the same amount of tracks as signal and other channels.

The input observables are mostly geometrical variables that give a measure of how well the two tracks form a vertex, how much they are separated from the PV, or how they point to it. The input variables are listed below.

Input variables

- $\boldsymbol{\chi^2_{\text{vertex}}}$ is the χ^2 of the vertex made up by the electron and muon candidates.
- **DOCA** is the distance of closest approach between the electron and muon candidates.
- **min(IP)** is the minimum of the impact parameter, *i.e.* distance of closest approach to the best PV, of the muon and electron candidates.
- $\boldsymbol{\text{IP}(B^0_{(d/s)})}$ is the impact parameter of the line given by the vertex and direction of the momentum of the $B^0_{(d/s)}$ candidate with respect to its best PV.
- $\boldsymbol{\chi^2_{\text{IP}}(B^0_{(d/s)})}$ is the difference in χ^2 of including the line given by the vertex and the direction of the momentum of the $B^0_{(d/s)}$ candidates as a track in the PV vertex fit χ^2 or not.
- **DIRA** is the cosine of the direction angle, *i.e.* the angle between the momentum direction and the flight direction of the $B^0_{(d/s)}$ candidates. The latter is given by the vector pointing from the PV to the vertex of the $B^0_{(d/s)}$ candidates.

The BDT is trained using simulation of $B^0_s \to \mu^+\mu^-$ decays as signal and simulation of inclusive $b\bar{b} \to \mu\mu X$ as background sample. The candidates used for the training have

to pass a selection similar to the aforementioned signal selection without the particle-identification. For the full selection and more details regarding the BDTS classifier see [57].

In the end, the distribution of the BDTS classifier is transformed to be uniform between 0 and 1 for the signal training sample. To achieve an efficiency of about 95%, the requirement for all channels is set that the classifier output has to be higher than 0.05.

3.3.3 BDT

The most separating classifier in the form of a BDT, hereafter colloquially called BDT, is constructed by adding input variables that are specifically suited for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$. The signal sample for training is simulated $B_s^0 \to e^\pm\mu^\mp$ events. The background sample consists of data with the same selection as for $B^0_{(d/s)} \to e^\pm\mu^\mp$, but where the muon and the electron have the same sign, such that it for sure contains only background. It is cross-checked that the variables have the same distribution as the opposite-sign signal sample by comparing it with data far away from the B^0 and B_s^0 masses to avoid potential biases.

With respect to the BDTS, more geometrical and also kinematic variables are added. An important addition is the use of isolation variables. They essentially give a measure how many other tracks are around them. As combinatorial backgrounds come from separate b-hadron decays, it is likely the electron and muon candidates have additional tracks around them, whereas signal would not. Special care has to be taken with isolation variables, since there are often quite significant differences between simulation and data. This is not a problem, as the response of the BDT is calibrated on data, as will be discussed in section 5.1.2. For the BDTS this is not necessary and it is therefore calibrated with the rest of the kinematic selection based on square cuts, discussed in section 4.1.1.

The input variables of the BDT are listed below.

Input variables

- $\sqrt{\textbf{min}(\textbf{IP}_{\chi^2})}$ is the square root of the minimum of the χ^2_{IP} of the electron and muon candidates. The square root is just used for computational reasons.
- $\textbf{IP}(\boldsymbol{B^0_{(d/s)}})$ is the same as used for the BDTS classifier.
- **DOCA** is the same as used for the BDTS classifier.
- $\textbf{FD}(\boldsymbol{B^0_{(d/s)}})$ is the flight distance, *i.e.* the distance of the decay vertex made up by the electron and muon candidates and the best PV.
- $\boldsymbol{\chi^2_{\text{DV}}}$ is the flight distance significance, constructed by the difference between adding the vertex in the PV vertex fit or not.
- $\textbf{cos}\,\boldsymbol{nk}$ is the cosine of the angle between the momentum of the muon candidates in the rest-frame of the B candidates and the vector perpendicular to the B momentum and z-axis in the lab frame. This angle is sensitive to the polarisation of the B candidates [58]. Since B mesons are pseudo-scalars, they are not polarised and therefore have a flat distribution in $\cos nk$ [59]. Due to a strong efficiency dependence on $\cos nk$, this looks more like a distribution proportional to $1 + (\cos nk)^2$. For

background, the distribution is peaking at zero due to the strong angular requirements on particles of separate decays to form a vertex.

- $\boldsymbol{\Delta\eta}$ is the difference in pseudo-rapidity between the electron and muon candidate.
- $\boldsymbol{p_{\mathrm{T}}(B^0_{(d/s)})}$ is the transverse momentum of the $B^0_{(d/s)}$ candidate.
- $\boldsymbol{p_{\mathrm{T,max}}}$ is the maximum of the transverse momentum of the electron and muon candidate.
- $\boldsymbol{I(B^0_{(d/s)})}$ is the isolation variable defined by,

$$I(B^0_{(d/s)}) = \frac{p_{\mathrm{T}}(B^0_{(d/s)})}{p_{\mathrm{T}}(B^0_{(d/s)}) + \sum_{\text{other tracks}} p_{\mathrm{T}}} \ , \tag{3.2}$$

  where the tracks that are summed over are within a cone of $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 1$ from the $B^0_{(d/s)}$ candidate direction.

- $\boldsymbol{I(e\mu)}$ is the sum of the daughter track isolations. Track isolation is defined by the number of tracks forming a good vertex with the track, excluding the signal tracks. What constitutes a good vertex is that the signal candidate track and the other track pass the requirement that,

$$\frac{|\vec{p}^{\ \text{track}} + \vec{p}^{\ \text{other}}| \sin\alpha}{|\vec{p}^{\ \text{track}} + \vec{p}^{\ \text{other}}| \sin\alpha + p_{\mathrm{T}}^{\ \text{track}} + p_{\mathrm{T}}^{\ \text{other}}} < 0.6 \ , \tag{3.3}$$

  where α is the angle between the momentum of the sum of the track and other track with respect to line made up by the vertex and the PV.

- $\boldsymbol{\tau_{B^0_s}}$ is the proper decay time of the $B^0_{(d/s)}$ candidate, assuming the B^0_s mass.

Their distributions, both for the signal and the background training sample, are shown in figs. 3.5 and 3.6.

It is preferable that the BDT is independent of the invariant mass of $e^\pm\mu^\mp$, $m_{e^\pm\mu^\mp}$. This will make the likelihood fit, discussed later in chapter 5, more easy. The correlation between the BDT and $m_{e^\pm\mu^\mp}$ is about 4% and hence sufficiently small. Further checks have been made to ensure that the BDT is not over-trained.

The BDT output is, just like the BDTS, transformed to be uniform between 0 and 1 for the signal training sample. The BDT is used to categorise the dataset in bins of the classifier output. A requirement is set that the BDT is larger than 0.25, as it later turned out that in the fit the bin $[0, 0, 25]$ did not add any sensitivity. It should be noted that the determination of the efficiency of this requirement is not calculated in chapter 4, but seen as part of the PDF of the BDT and subsequently discussed in section 5.1.2.

The full set of multivariate requirements for $B^0 \to e^\pm\mu^\mp$ and $B^0_s \to e^\pm\mu^\mp$ are listed in table 3.11. This concludes all the selection done on the signal dataset. The selection results in 476 $B^0_{(d/s)} \to e^\pm\mu^\mp$ candidates in the full dataset of 2011 and 2012.

Figure 3.5: First set of distributions of input variables of the BDT for signal and background training samples for the BDT. Signal is $B_s^0 \to e^\pm\mu^\mp$ simulation and background is wrong-sign $e^\pm\mu^\pm$ data. Rest of the input variables are shown in fig. 3.6.

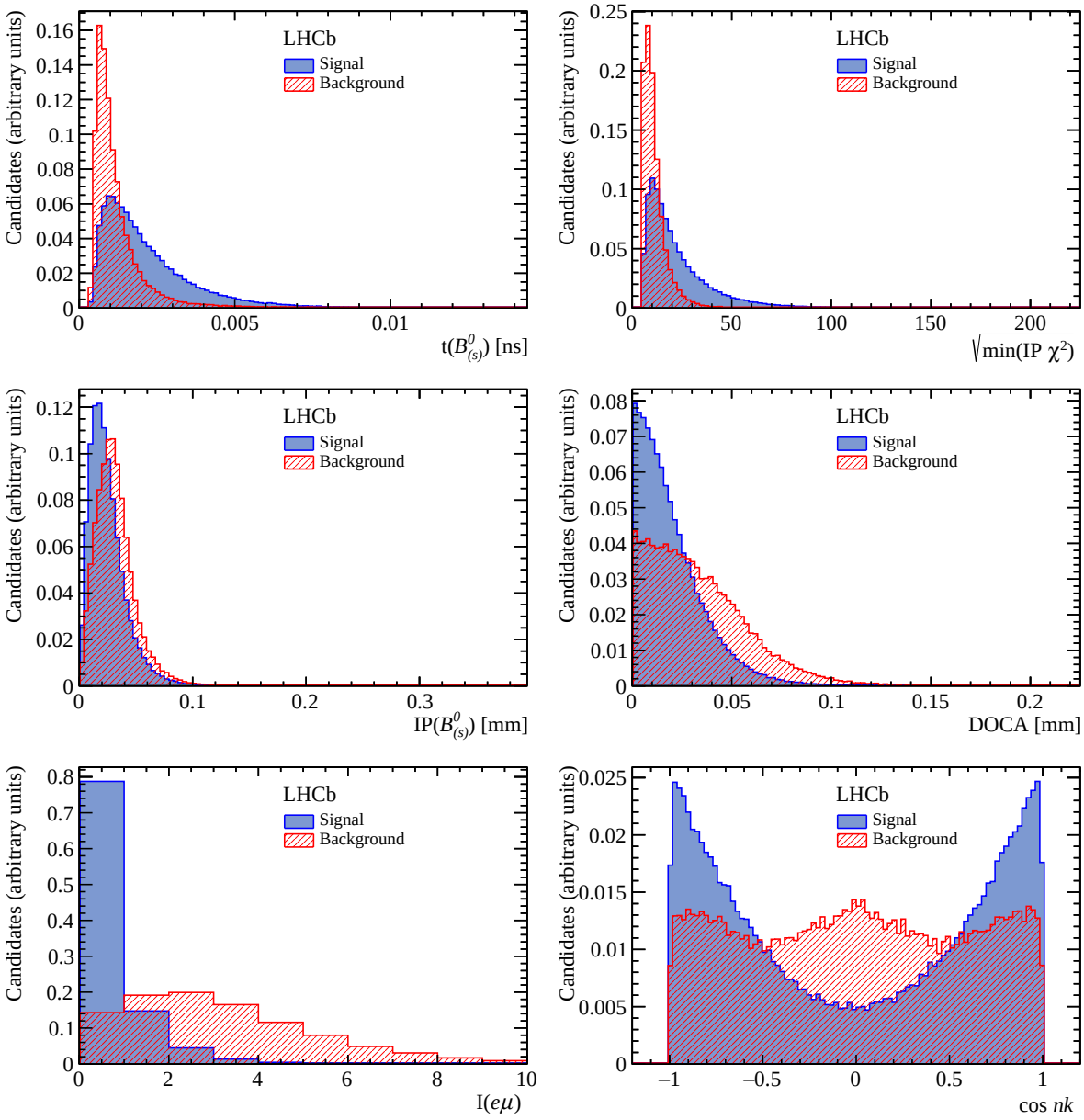
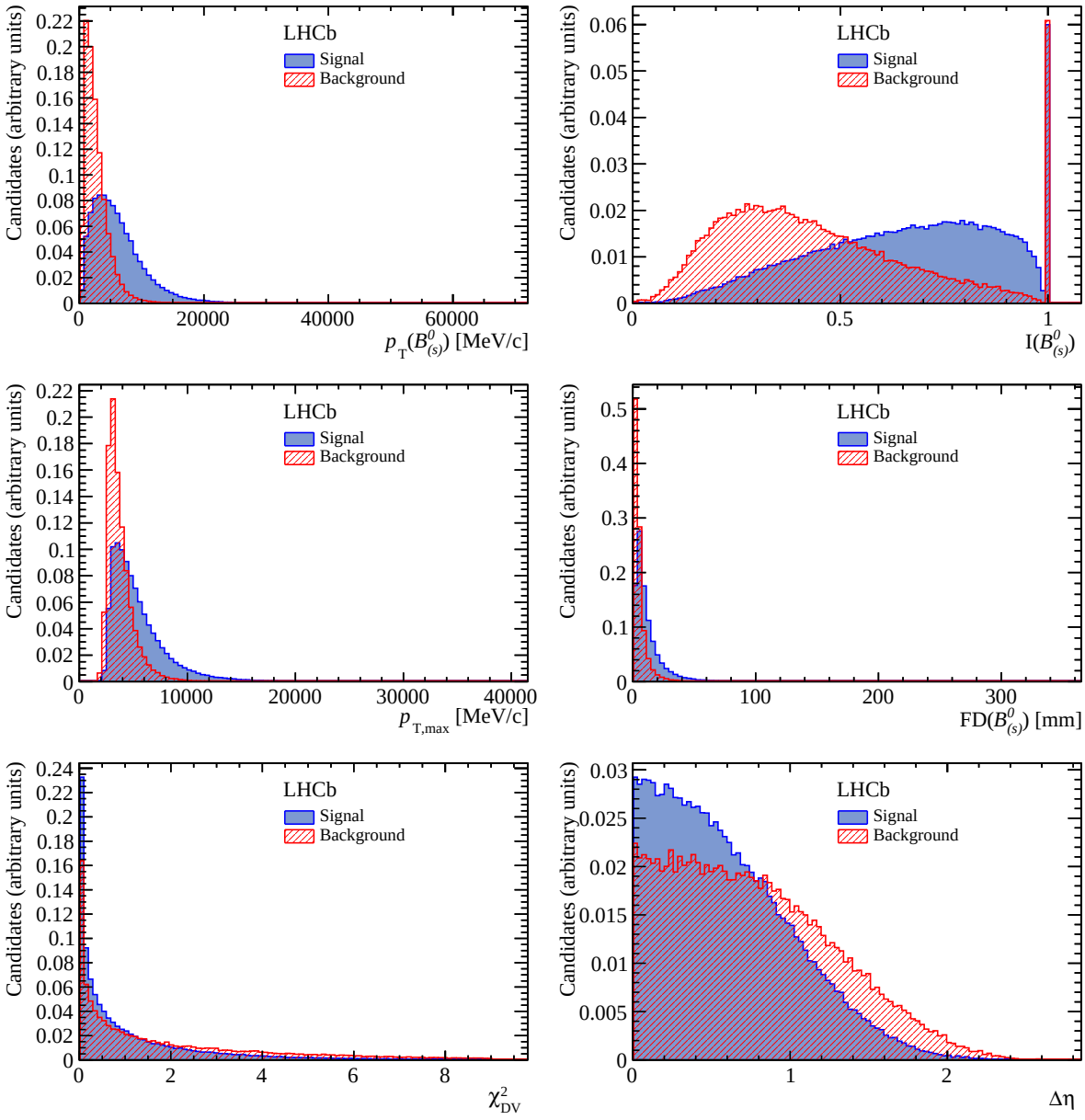


Table 3.11: Multivariate classifier requirements for $B^0_{(d/s)} \to e^\pm\mu^\mp$ candidates.

Candidate	Variable	Value
$B^0_{(d/s)}$	BDTS	> 0.05
	BDT	> 0.25

Figure 3.6: Second set of distributions of input variables of the BDT for signal and background training sample for the BDT. Signal is $B_s^0 \to e^{\pm}\mu^{\mp}$ simulation and background is wrong-sign $e^{\pm}\mu^{\pm}$ data. Rest of the input variables are shown in fig. 3.5.



Chapter 4

Normalisation

To obtain branching fractions of $B^0 \to e^\pm \mu^\mp$ and $B_s^0 \to e^\pm \mu^\mp$, their respective yields are normalised to well-known b-hadron decays. The normalisation procedure is discussed in this chapter, while the signal-yield extraction is discussed in chapter 5.

The normalisation procedure requires correcting the yields of the signal and of the normalisation channels by their selection efficiencies. In case the signal and normalisation channels are not the decay of the same type of b-hadron, a further correction by their respective b-hadron production fraction has to be applied. Taking all this into account, the signal branching fraction can be determined as follows:

$$\mathcal{B}_{\text{sig}} = \frac{\mathcal{B}_{\text{norm}}}{N_{\text{norm}}} \frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}} \frac{f_{\text{norm}}}{f_{\text{sig}}} \cdot N_{\text{sig}} \ , \tag{4.1}$$

where $\mathcal{B}$ denotes the branching fraction, N the yield, ε the total efficiency and f the b-hadron production fraction; the signal and normalisation modes are indicated by the subscript 'sig' or 'norm' respectively.

The choice of normalisation channel is a trade off between the precision with which the branching fraction of the normalisation channel is known and the precision with which the ratio of efficiencies can be determined. The latter is often the main concern, as a careful choice can lead to cancellations of uncertainties due to similarities between the signal and the normalisation mode. With this in mind, two modes are chosen. One, the decay of $B^0 \to K^+\pi^-$, is chosen because of the similarity in selection with the multivariate classifier (two-body final state). The other, the decay of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, is chosen because of similarities in trigger and particle-identification selection (muons in the final state). A weighted sum of the two normalisation factors is used. The determination of the efficiencies of both the signal and normalisation modes are discussed in section 4.1.

As the b-hadron species of the signal are B^0 and B_s^0, and for the normalisation channels B^+ and B^0, the yields have to be corrected by their respective production fractions. No decays channels of B_s^0 are used as normalisation channel, as no absolute branching fractions have been precisely measured for B_s^0 [59]. The b-hadron production fractions used in this analysis are obtained from dedicated LHCb analyses [60].

Next, the determination of the yields of the two normalisation modes is discussed: the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ decay channel in section 4.2 and section 4.3, respectively. The signal yield extraction is discussed in chapter 5. In the last part of the chapter, section 4.4, the normalisation factors are combined and discussed, including a cross-check of the efficiency determination of the normalisation modes: the ratio of

efficiency-corrected yields of the two channels is compared to the expected value determined with their branching fractions from the PDG [59].

4.1 Efficiencies

For this analysis, the determination of the efficiencies of reconstructing and selecting a certain decay is based on simulations. Every part of the data flow with each of its corresponding selection has a specific simulation. The efficiencies are split into stages, each conditional on the previous stage, such that each has a well-defined dedicated determination, either with or without data-driven corrections to take into account possible differences between data and simulation.

Following the selection described in chapter 3, the factorised efficiency for signal is given by,

$$\epsilon = \epsilon_{\text{gen}} \cdot \epsilon_{\text{sel|gen}} \cdot \epsilon_{\text{trig|sel}} \cdot \epsilon_{\text{PID|trig}} \ , \tag{4.2}$$

where ϵ_{gen} is the generation efficiency of the simulation, $\epsilon_{\text{sel|gen}}$ is the reconstruction and selection efficiency without any particle-identification and trigger requirements, $\epsilon_{\text{trig|sel}}$ is the trigger efficiency, and $\epsilon_{\text{PID|trig}}$ is the particle-identification efficiency. These efficiencies are discussed in sections 4.1.1 to 4.1.3, respectively. The sequence of conditions is chosen in such a way that it makes the efficiency determination the most reliable and easy to determine. For normalisation channels the choices may differ slightly; whenever this is the case, it is explicitly mentioned. In general, the ordering choices will be elaborated in the coming sections.

For the signal modes, only $B_s^0 \to e^\pm\mu^\mp$ simulation is used. For the $B^0 \to e^\pm\mu^\mp$ case, the small corrections due to the slightly smaller mass will be discussed in section 4.4. In this simulation, an even mixture of heavy and light eigenstates of B_s^0 decaying to $e^\pm\mu^\mp$ is used. Since a priori it is not known if the hypothetical decay only occurs through the *CP*-even, odd or a mixture of eigenstates, different effective lifetimes of the B_s^0 have to be considered. The extreme cases are considered and efficiencies are calculated for each case by reweighing the lifetime distribution in simulation to the desired effective lifetime. For the sake of brevity, only the results for the default simulation will be presented. Since the B^0 meson does not have a significant lifetime difference between the two mass eigenstates, no such consideration is needed as for the B_s^0 meson.

4.1.1 Generation, reconstruction and kinematic selection efficiencies

In simulation, only those events are stored that generate final-state particles of the decay in question that are within the geometrical acceptance of LHCb. This is done to save disk space and CPU time. In practice, it is a requirement on the pseudo-rapidity of the final-state particles. The efficiencies are calculated by keeping track of how many events are generated and discarded. The resulting efficiencies are reported for signal and normalisation channels in table 4.1. Notable are the similarities between the $B_s^0 \to e^\pm\mu^\mp$ and $B^0 \to K^+\pi^-$ modes, as they have both two charged (pseudo) stable particles as final state. The efficiency of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is lower as an additional particle has to be within the acceptance of LHCb. The reported uncertainties are statistical and obtained with the sample size of the simulation. No systematic uncertainty is assigned

as the resolution of the pseudo-rapidity is far more precise than the wide margins of the selection.

Table 4.1: Generation efficiencies for signal and normalisation channels in simulation. Essentially it is a selection based on the pseudo-rapidity of the final-state particles. Uncertainties are statistical, dictated by the sample size of the simulation.

Simulation sample	ϵ_{gen}
$B_s^0 \to e^\pm \mu^\mp$	18.56 ± 0.08 %
$B^0 \to K^+\pi^-$	18.97 ± 0.05 %
$B^+ \to J/\psi(\to \mu^+\mu^-)K^+$	16.63 ± 0.07 %

After generation, the final-state particles are propagated through the detector, simulating each part of the detector it goes through, eventually generating hits in the detector, analogous to what happens in data. From this point, the same software used for data can be applied to reconstruct and select decays. By keeping track of the links between generated particles and simulated hits, reconstructed objects can be linked to the particle that is originally generated. The calculation of efficiencies is performed by comparing the number of reconstructed objects that pass the selection and have the right number of links with the total number of generated particles. The resulting efficiencies of the reconstruction and selection without trigger and full particle-identification requirements are reported in table 4.2. The only particle-identification requirements that are included in this selection is `isMuon`, both for $B_s^0 \to e^\pm\mu^\mp$ and $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, and a highly-efficient requirement, about 99%, on the `DLLe` to be larger than -2 for the electron. The inefficiency and possible resulting systematic uncertainty is negligible compared to the final particle-identification requirement set on $B_s^0 \to e^\pm\mu^\mp$ that is much harsher. In addition, `isMuon` is only a matching of a track in the muon stations to the track in the main tracking stations. Subsequently, it is similarly well-reproduced in simulation as for the rest of reconstruction and selection. The reported uncertainties are statistical, corresponding to the simulation sample size. Systematic uncertainties are larger, but are determined by reweighing certain distributions of variables affected by data-simulation differences to its distributions obtained from data. To not double count uncertainties, this is calculated once for the efficiency of the final trigger stage, *i.e.* `HLT2`. As for the generation efficiencies, note the similarities of the efficiencies between signal and normalisation mode.

Table 4.2: Combined reconstruction and kinematic selection efficiencies for signal and normalisation channels in simulation. Uncertainties are statistical, obtained using the sample size of the simulation. It includes the loose selection on the general BDT, BDTS, on all channels. For $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, it also includes the efficiency of the `isMuon` requirements on the two muons, the only particle-identification requirement for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$.

Simulation sample	$\epsilon_{\text{sel}\|\text{gen}}$
$B_s^0 \to e^\pm \mu^\mp$	$23.74 \pm 0.04\%$
$B^0 \to K^+\pi^-$	$23.15 \pm 0.06\%$
$B^+ \to J/\psi(\to \mu^+\mu^-)K^+$	$18.66 \pm 0.02\%$

4.1.2 Trigger

Determining the efficiencies of the trigger of LHCb poses a two-fold challenge. First, the hardware stage is less well reproduced in simulation than the offline reconstruction and selection steps. Second, the thresholds for L0 (and to a lesser extend HLT1) depend on the conditions of the LHC during the year of data taking. For simulation, the most representative configuration is chosen, not the full set of configurations. For these reasons, a data-driven technique is chosen to obtain the efficiencies of the L0 and HLT1 stages: the TISTOS method [61].

This method is based on determining efficiencies on a sample that is independently triggered, *i.e.* not triggered by the decay the efficiencies will be determined with. As mentioned in section 2.5, triggering independent of signal is called TIS and triggering on signal is called TOS. To determine whether a particle fired the trigger or not, a matching of the reconstructed signal particles is performed with the reconstructed objects on which the trigger decisions are based on. A trigger decision is TOS if there is a match and the matched reconstructed-object has the right properties to fire the trigger line by itself. A trigger decision is TIS if there is a reconstructed object that is able to fire the trigger line by itself, but does not match to the signal particles.

With the TISTOS method, the assumption is made that the efficiency of a TOS selection is sufficiently well approximated by,

$$\epsilon_{\texttt{TOS}} \approx \epsilon_{\texttt{TOS|TIS}} = \frac{\epsilon_{\texttt{TISTOS}}}{\epsilon_{\texttt{TIS}}} = \frac{N_{\texttt{TISTOS}}}{N_{\texttt{TIS}}} \ , \tag{4.3}$$

where $N_{\texttt{TISTOS}}$ and $N_{\texttt{TIS}}$ are the yield of the calibration channel with TIS and TOS and TIS-only selection respectively. This approximation holds if the TIS selection is indeed independent from, *i.e.* uncorrelated to, TOS.

The hard part of the TISTOS method is to actually have an independent trigger. A choice one could make is to use a general collection of triggers on particles coming from another PV. For the Run 1 luminosities this is a very limited sample. Therefore, it is chosen to apply such a selection to the whole event, but excluding the signal particles. Most triggers are designed to trigger on b-hadron decays. Subsequently, triggering on the whole of the rest of the event means often triggering on a b-hadron decay coming from a b-quark related to the b-quark of the signal. This is due to the fact that most b-quarks originate from $b\bar{b}$-pairs produced by gluon-gluon fusion. This means that such a TIS selection is quasi independent, but only correlated through the kinematics of the b-hadron used as signal. This problem can be alleviated by determining these efficiencies in bins depending on the b-hadron kinematics. Subsequently, one can obtain a proper efficiency by reweighing to the right distribution of its kinematics. Beside for correcting for the effect of the kinematics of the b-hadron, introduced by triggering on kinematic thresholds like the p_{T} of the muon, reweighing is also needed to translate the efficiency from the kinematic distribution of the calibration channel to the one of the signal channel in the first place.

As shown in section 3.1, the signal uses a trigger for $\texttt{L0} \otimes \texttt{HLT1}$ either on the muon or the electron. Therefore, two samples are needed for the TISTOS method. One for the triggering on a single muon, the other on a single electron.

For the muon, the TISTOS method is applied on the dataset of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$. The TIS selection comprises L0Muon, to ensure the same SPD hit multiplicity requirements

Figure 4.1: Binning in muon p_{T} en IP with the distribution for $B_s^0 \to e^{\pm}\mu^{\mp}$ overlaid (left) and efficiencies in those bins (right) obtained with the TISTOS method from the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ sample. The binning is chosen to have most bins in regions where the efficiency changes most.

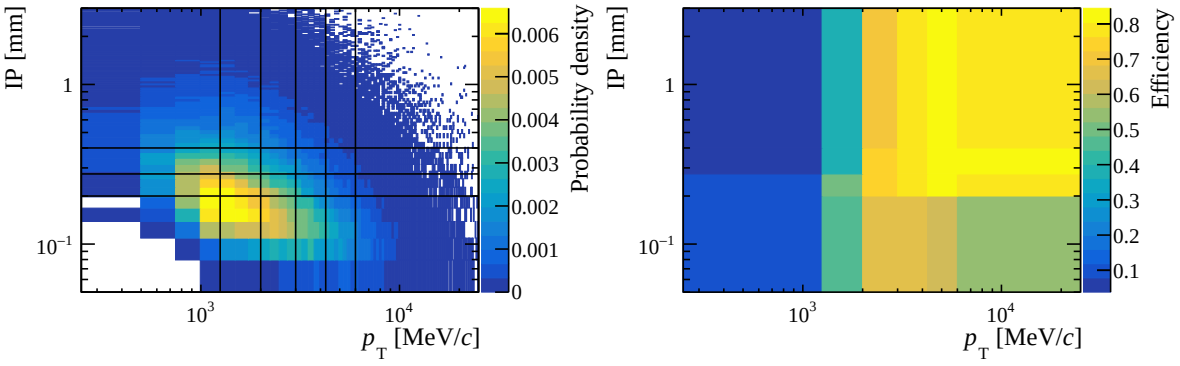
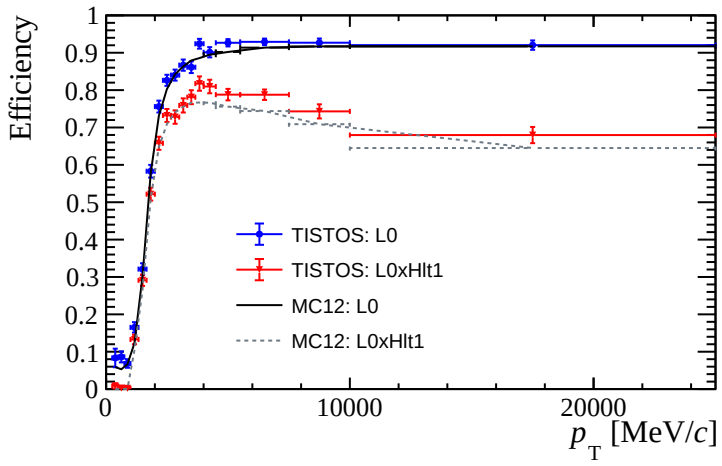


Figure 4.2: Efficiencies, both for `L0` and `L0`$\otimes$`HLT1`, depending on muon p_{T} obtained with TISTOS method. The p_{T} is the strongest requirement and this shows clearly the threshold set by `L0Muon` and how similar data and simulation are.



as for signal, and general triggers for `HLT1` and `HLT2`. A similar yield extraction as for the normalisation with this channel is used, which will be described in section 4.2. The efficiencies are determined binned in p_{T} and IP of the muon. These variables are chosen since `L0Muon` sets requirements on the p_{T} of the muon and `HLT1` also sets requirements on IP related variables. This binning ensures that the efficiency can be reweighed to have the proper kinematics of the signal or normalisation channel. The two-dimensional binning with the distribution for $B_s^0 \to e^{\pm}\mu^{\mp}$ and the resulting efficiencies of this binning from the TISTOS method are shown in fig. 4.1. As the p_{T} is the strongest requirement, the dependence of the efficiency only on p_{T} is shown in fig. 4.2. It also shows the degree of how well the simulation reproduces the data for both `L0` and `L0` $\otimes$ `HLT1` selections. Although there are some significant differences, it stays within a few percent.

For the electron, the TISTOS method is applied on a dataset of $B^+ \to J/\psi(\to e^+e^-)K^+$. The `TIS` selection is the same as for the muon case. The only difference with respect to the muon case is that `L0Electron` sets requirements on the transverse energy deposit,

Figure 4.3: Binning in electron E_{T} en IP with the distribution for $B_s^0 \to e^{\pm}\mu^{\mp}$ overlaid (left) and efficiencies in those bins (right) obtained with TISTOS method from the $B^+ \to J/\psi(\to e^+e^-)K^+$ sample. The binning is chosen to have most bins in regions where the efficiency changes most.

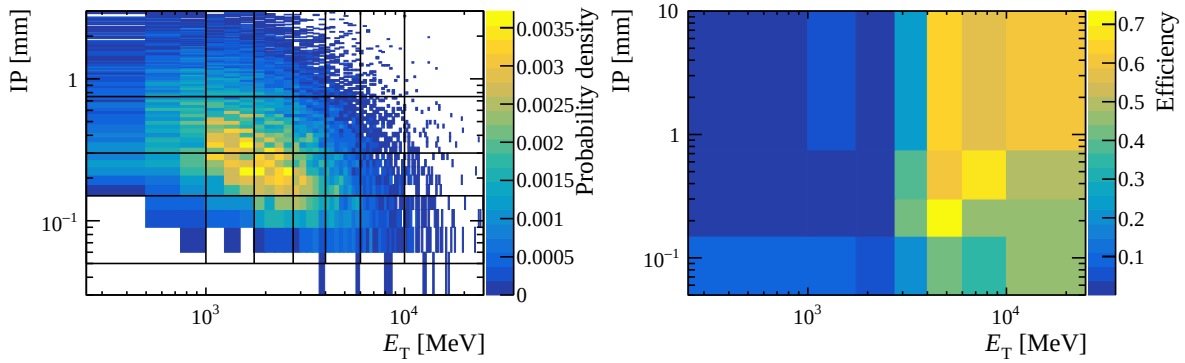
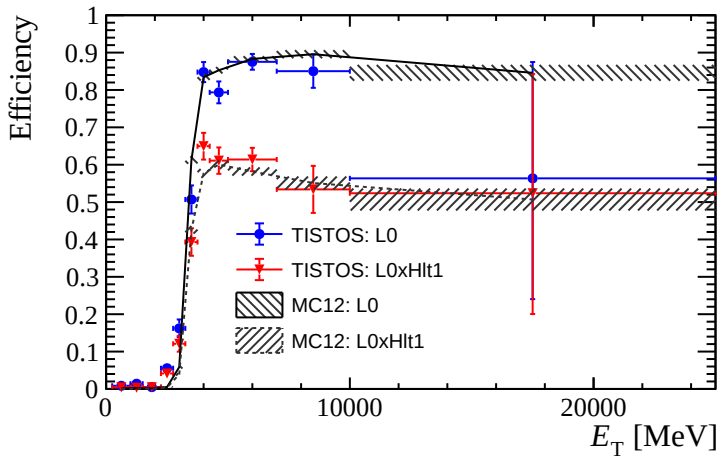


Figure 4.4: Efficiencies, both for `L0` and `L0` $\otimes$ `HLT1`, depending on electron E_{T} obtained with TISTOS method. The E_{T} is the strongest requirement and this shows clearly the threshold set by `L0Electron`.

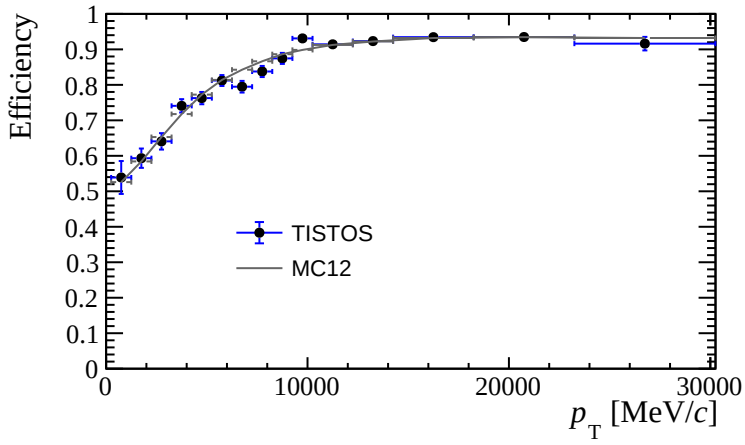


E_{T}, in the electromagnetic calorimeter. For this reason the variables used for reweighing are E_{T} and IP. There are no significant differences observed, albeit that the statistical uncertainties are larger than for the muon case.

The `L0` $\otimes$ `HLT1` $\otimes$ `HLT2` efficiencies for $B_s^0 \to e^{\pm}\mu^{\mp}$ are calculated by combining the `L0` $\otimes$ `HLT1` efficiencies for the single muon and electron, binned in IP and p_{T} or E_{T} of the muon and electron respectively, multiplying with the `HLT2` efficiency in that bin. The `HLT2` efficiency is obtained from simulation. The resulting `L0` $\otimes$ `HLT1` $\otimes$ `HLT2` efficiencies in these bins are weighted according to the respective distributions for $B_s^0 \to e^{\pm}\mu^{\mp}$ in simulation. Systematic uncertainties on this procedure are obtained by varying the binning schemes and reweighing the p_{T} of the B to the efficiency-corrected distribution obtained from $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and the amount of hits in the SPD as occupancy variable. The resulting efficiencies are reported in table 4.3.

The efficiency of the full trigger selection for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is determined using the TISTOS method with $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ itself. This results in a small and therefore negligible correlation, because `TIS` efficiencies are small and subsequently

Figure 4.5: Efficiency of L0 ⊗ HLT1 ⊗ HLT2 for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ determined with the TISTOS method with $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ itself for 2012 versus the p_{T} of the B. The latter is done to alleviate biases caused by the TIS selection. The gray band denotes the efficiency in simulation. Data and simulation agree well within uncertainties.



the overlap with a TOS selection is small. The obtained L0 ⊗ HLT1 ⊗ HLT2 efficiencies for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ in bins of p_{T} of the B for 2012 are shown in fig. 4.5 and are in agreement with the 2012 simulation. As mentioned before, the only bias that has to be corrected for is that a TIS selection biases the B hadron kinematics. Accordingly, a reweighing of the efficiencies in bins of the p_{T} of the B is applied. For the reweighing, the distribution of the p_{T} of the B from data is used. Since one has to reweigh to the p_{T} distribution before the trigger, a correction has to be applied. Using the trigger efficiencies binned in p_{T}, one can unfold the p_{T} distribution before the trigger by multiplying the distribution after the trigger with the inverse of the efficiency. The distribution of p_{T} of the B is selected by setting a strong requirement on the invariant mass around the B peak. The fraction of combinatorial in this sample is about 0.3% and gives a negligible contribution. Systematic uncertainties due to this determination are calculated by varying the binning of the p_{T} and the invariant-mass requirement. The reweighed L0⊗HLT1⊗HLT2 efficiencies of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ are reported in table 4.3.

For $B^0 \to K^+\pi^-$, the same method as for $B_s^0 \to e^\pm\mu^\mp$ is applied, except that L0 selects solely on L0Hadron and the calibration channel is $B^0 \to K^+\pi^-$ itself. Its resulting efficiency is reported in table 4.3.

4.1.3 Particle identification

For the determination of the particle-identification efficiencies, an approach similar to the trigger efficiencies is used. Just like the trigger efficiencies, the particle-identification variables are not simulated accurately enough. This is mostly due to occupancy differences between simulation and data. As an example of why in simulation occupancy is different than in data, consider that only particles with a certain minimum amount of momentum are propagated through the detector, to reduce CPU time. Many secondaries, *i.e.* particle produced by material interactions, are therefore not simulated, resulting in lower occupancies. To illustrate the effect this has on particle identification, take for example

Table 4.3: Efficiencies of the trigger selection, L0 ⊗ HLT1 ⊗ HLT2, for signal and normalisation channels. They are determined with a combination of the data-driven TISTOS method and simulation. For $B^0 \to K^+\pi^-$, the selection includes particle-identification, which will be discussed in section 4.1.3.

Channel	$\epsilon_{\text{trig\|sel}}$	
$B_s^0 \to e^\pm\mu^\mp$	!HasBremAdded	HasBremAdded
	$72.6 \pm 0.2\,(\text{stat}) \pm 1.5\,(\text{syst})\%$	$62.1 \pm 0.2\,(\text{stat}) \pm 1.5\,(\text{syst})\%$
$B^0 \to K^+\pi^-$	$21.23 \pm 0.19\%$	
$B^+ \to J/\psi(\to \mu^+\mu^-)K^+$	$75.8 \pm 0.6\,(\text{stat}) \pm 1.0\,(\text{syst})\%$	

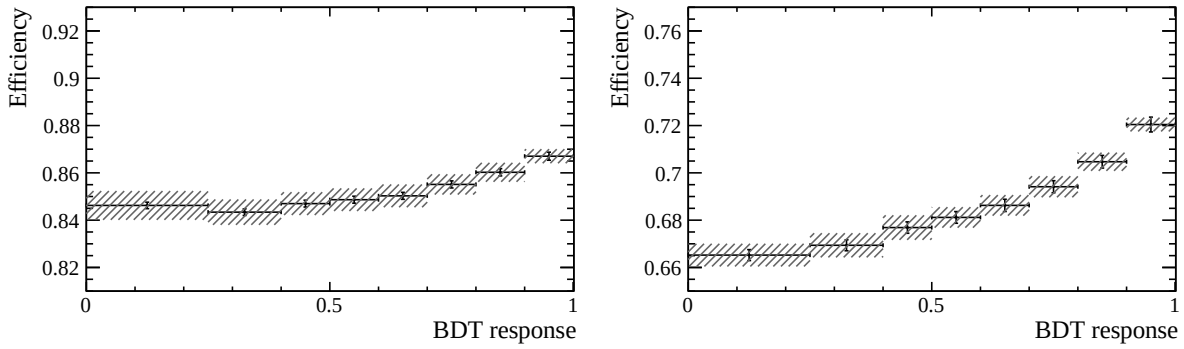
the likelihood ratios obtained from the two RICH detectors. As the event gets busier, more photons unrelated to particle candidates will dilute the likelihood ratios. Similar arguments hold for the calorimeter. For these reasons, a data-driven approach is needed to calculate the particle-identification efficiencies.

At LHCb, a centrally-maintained calibration tool is available, called the PIDCalib package [62]. It is based on a set of decay channels, where through tight requirements on a subset of the final-state particles (called the tag) the identity of the other particle or particles (called the probe) can be deduced. No particle-identification requirements are set on the probe. This technique is called tag-and-probe. The decay channels used by PIDCalib that are relevant for this analysis are the decays of $J/\psi \to \mu^+\mu^-$ coming from b-hadron decays, $B^+ \to J/\psi(\to e^+e^-)K^+$ and $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ for the muon, electron and pion or kaon respectively. Invariant-mass likelikelihood fits have been performed beforehand to obtain per-candidate weights to subtract backgrounds [63]. Efficiencies are subsequently obtained by determining the background-subtracted yields before and after setting particle-identification requirements on the probe particle or particles. This procedure is performed in bins of variables that need reweighing. Mostly, these are variables used to correct kinematics from calibration to signal channel, like the p_T and η of the relevant particle or a variable representing the occupancy of the event.

For $B_s^0 \to e^\pm\mu^\mp$, for both the muon and electron calibration sample, a reweighing of the obtained efficiency is performed in p_T and η of the track. For the electron, also the distribution of the number of hits in the SPD has been reweighed, as the electron is more affected by occupancy than the muon. For example, for electrons, the information from the reconstruction of bremsstrahlung photons adds to the particle-identification and the higher the occupancy, the more likely random photons are associated as bremsstrahlung photons. Note that the occupancy differences between signal and calibration channel are not large. It is mostly the occupancy difference between simulation and data that is large. For the reweighing of the number of hits in the SPD, the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ dataset is used as proxy, similar to the procedure of the trigger efficiency determination of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$. For the reweighing of p_T and η of the track, simulation of $B_s^0 \to e^\pm\mu^\mp$ is used. As explained in the beginning of section 4.1, the particle-identification efficiencies have to be determined per BDT bin and are reported in fig. 4.6.

For $B^0 \to K^+\pi^-$, the kaon and pion particle-identification efficiencies are determined by reweighing the binned efficiencies from PIDCalib in p and η of the tracks. For $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, the only particle-identification requirement is isMuon. As

Figure 4.6: Particle-identification efficiencies for $B_s^0 \to e^\pm\mu^\mp$ depending on BDT classifier and if bremsstrahlung has been reconstructed (left) or not (right) for 2012. Note the difference in overall efficiency for the latter category. Black vertical lines denote statistical uncertainty, grey bands denote systematic uncertainties.



explained in section 4.1.1, this is included in the general selection and reconstruction efficiency determined with simulation. The efficiency of requiring `isMuon` on both the muons from $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is about 98%. On the kaon, no particle-identification requirement is set.

The systematic uncertainties of the `PIDCalib` procedure for $B_s^0 \to e^\pm\mu^\mp$ and $B^0 \to K^+\pi^-$ are obtained by varying the binning schemes. The efficiencies are reported in table 4.4. For $B^0 \to K^+\pi^-$, the systematic uncertainty is not used at this stage and therefore not quoted to avoid double counting. These are incorporated in the data-driven determination of the PDF of the BDT classifier with $B^0 \to K^+\pi^-$ in section 5.1.2.

Table 4.4: Efficiencies of the particle-identification requirement for signal and normalisation channels determined with the `PIDCalib` tool [62]. For $B_s^0 \to e^\pm\mu^\mp$, the efficiencies are conditional on the trigger selection, the rest on the general selection. For $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, the efficiency of the `isMuon` requirement is calculated with simulation and is part of the efficiency reported in table 4.2. The equivalent calculated with `PIDCalib` is reported here. They are in agreement within uncertainties.

Channel	ϵ_{PID}	
$B_s^0 \to e^\pm\mu^\mp$	`!HasBremAdded`	`HasBremAdded`
	$69.47 \pm 0.16\,(\text{stat}) \pm 0.37\,(\text{syst})\%$	$85.43 \pm 0.09\,(\text{stat}) \pm 0.37\,(\text{syst})\%$
$B^0 \to K^+\pi^-$	$38.50 \pm 0.01\,(\text{stat})\%$	
$B^+ \to J/\psi(\to \mu^+\mu^-)K^+$	$97.81 \pm 0.02\,(\text{stat})\%$	

4.2 $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$

The yield of the normalisation channel $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is determined from the distribution of its most distinguishing observable, the invariant mass of the $\mu^+\mu^-K^+$ combination, denoted by $m_{J/\psi K^\pm}$, where $\mu^+\mu^-$ is constrained to the J/ψ mass using the

`DecayTreeFitter` algorithm [64]. Subsequently, the yields of the components in the distribution of $m_{J/\psi K^\pm}$ are extracted with an extended-likelihood maximisation [59, 65].

For signal, the observable $m_{J/\psi K^\pm}$ should be the mass of the B^+ meson, about $5279\,\text{MeV}/c^2$, neglecting the small corrections of final-state radiation. Therefore, the distribution is mostly given by the resolution of the invariant-mass, which is in itself not constant. At LHCb, it is common to describe the resulting distribution with the empirical double-sided Crystal Ball function, defined in eq. (A.1). Nonetheless, at high statistics, this description is not sufficient to account for the invariant-mass resolution. To overcome this problem, an analytical function, called the Hypatia function, has been developed to adapt the Crystal Ball function by convoluting the width with an analytical description of the distribution of the resolution [66]. Its definition is given in appendix A.2. The Hypatia function is used for the signal distribution in this likelihood fit. The mean and width of the core Gaussian of the Hypatia are left free in the fit, the other parameters are left free in the fit as well, but constrained with a Gaussian prior to values, including uncertainties, determined from a fit to simulation of $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+$ [67].

Aside from the signal component, the distribution of $m_{J/\psi K^\pm}$ also contains backgrounds. The main background is from particles that are, at least partly, unrelated to each other, *e.g.* particles that originate from other decays. This type is called combinatorial background. As momentum distributions behave exponential at high momenta, the resulting invariant-mass distributions are exponential as well. For this reason, it is modelled using an exponential. The coefficient of the exponential is left free in the fit.

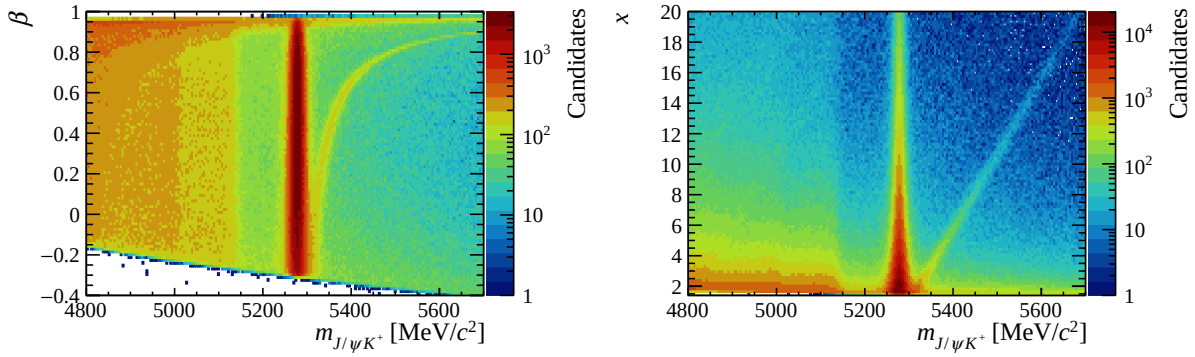
On top of combinatorial background, there are also backgrounds of which the final-state particles come from one decay of a b-hadron, but where the final-state particles are partly reconstructed, misidentified or both. In case of missing particles, one could for example miss a pion or a kaon. The resulting invariant mass would be lower than the b-hadron mass minus the pion or kaon mass. With a window of $m_{J/\psi K^\pm} \in (5180, 5680)\,\text{MeV}/c^2$, no partially-reconstructed decays are observed within this selection. This is because abundant decays like $B^0 \to J/\psi\,(\to \mu^+\mu^-)K^+\pi^-$, with missing a π^-, or $B_s^0 \to J/\psi\,(\to \mu^+\mu^-)K^+K^-$, with missing a K^+, fall out this window, taking into account a margin for the invariant-mass resolution.

In case of misidentified particles, a contribution comes from $B^+ \to J/\psi\,(\to \mu^+\mu^-)\pi^+$, as no particle-identification requirement on the kaon was used. The reason for this is that the amount of $B^+ \to J/\psi\,(\to \mu^+\mu^-)\pi^+$ is only about 4% relative to $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+$, and it can be modelled well. In addition, the advantage is that a particle-identification efficiency needs not to be determined. The modelling of $B^+ \to J/\psi\,(\to \mu^+\mu^-)\pi^+$ can be done well, because the decay is kinematically very similar to $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+$ and therefore practically has the same invariant-mass resolution, given that one assumes the right mass hypothesis for either the K^+ or the π^+. The transformation from one hypothesis to the other is given by,

$$m_{J/\psi K^\pm} \simeq \sqrt{m^2_{J/\psi\pi^+} + (m_K^2 - m_\pi^2)\Big(1 + \frac{p_{J/\psi}}{p_K}\Big)} \approx m_{J/\psi\pi^+}\Big[1 + \frac{(m_K^2 - m_\pi^2)}{2m^2_{J/\psi\pi^+}}\Big(1 + \frac{p_{J/\psi}}{p_K}\Big)\Big]\,, \tag{4.4}$$

where $m_{J/\psi\pi^+}$ is the invariant-mass under the right hypothesis, $p_{J/\psi}$ and p_K are the absolute momenta of the J/ψ and kaon respectively; the expression is expanded to first order in m/p for each particle [68, 69]. The following dimensionless variables, momentum

Figure 4.7: Two-dimensional distributions of momentum asymmetry (left), β, and momentum fraction (right), x, versus $m_{J/\psi K^\pm}$ in data for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ candidates. The linear dependence of the $B^+ \to J/\psi(\to \mu^+\mu^-)\pi^+$ background with x can be clearly seen. In addition, it shows partially-reconstructed backgrounds, safely below the minimal mass requirement of $5180\,\mathrm{MeV}/c^2$. The requirement of $x < 20$, *i.e.* $\beta < 0.9$, can be seen to reduce partially-reconstructed and misidentified background in the left figure.



fraction and momentum asymmetry respectively, are introduced,

$$x \equiv 1 + \frac{p_{J/\psi}}{p_K} \qquad\qquad \beta \equiv \frac{p_{J/\psi} - p_K}{p_{J/\psi} + p_K} \ . \tag{4.5}$$

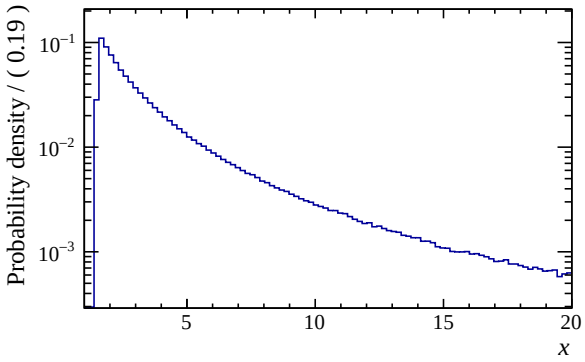
Both variables are useful to illustrate the behaviour of misidentified backgrounds. The advantage of β is that it has a fairly flat distribution, while for x the dependence on the shift in invariant mass is linear. The two two-dimensional distributions of $m_{J/\psi K^\pm}$ versus β and x are shown in fig. 4.7. The figure with $m_{J/\psi K^\pm}$ versus x shows the correctly identified signal peak of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and the misidentified $B^+ \to J/\psi(\to \mu^+\mu^-)\pi^+$ peak behaving constant and linearly with respect to x respectively. Aside from purely misidentified backgrounds, in the figure with $m_{J/\psi K^\pm}$ versus β, one can see that a part of the partially-reconstructed backgrounds, seen on the left side of both figures, moves into the signal window of $m_{J/\psi K^\pm} \in (5180, 5680)\,\mathrm{MeV}/c^2$ at high β due to misidentification. To reduce these backgrounds, the requirement of $x < 20$, *i.e.* $\beta < 0.9$, is set.

With eq. (4.4) and the distribution of x from $x_{\min}$ to $x_{\max}$, $f(x|\vec{\lambda})$, the PDF of $m_{J/\psi K^\pm}$ for $B^+ \to J/\psi(\to \mu^+\mu^-)\pi^+$ can be written as,

$$g^{\mathrm{misID}}(m_{J/\psi K^\pm}|\mu, \sigma, \vec{\lambda}) \propto \int_{x_{\min}}^{x_{\max}} dx \cdot g(m_{J/\psi K^\pm}|\widetilde{\mu}(\mu, x), \sigma) \cdot f(x|\vec{\lambda}) \ , \tag{4.6}$$

where $g(m_{J/\psi K^\pm}|\widetilde{\mu}(\mu, x), \sigma)$ is the PDF for the signal mode of the invariant-mass of which the used mass-hypotheses are correct; μ and σ are the mean and width of the signal mode respectively; $\widetilde{\mu}(\mu, x)$ is the transformation of eq. (4.4) and λ is the set of variables describing $f(x|\vec{\lambda})$. Note that eq. (4.6) can be generalised for any single-misidentified background. For modelling the distribution of the momentum fraction, shown in fig. 4.8, a power law is a good approximation for $x \gg 1$. The power law has a few nice properties. First of all, it is has one parameter, the power, and it can be approximated itself by a sum of exponentials [70]. This means that the integral from eq. (4.6) has an analytical solution if $g(m_{J/\psi K^\pm}|\widetilde{\mu}(\mu, x), \sigma)$ is a Gaussian distribution. For low values of x, an approximation

Figure 4.8: Distribution of the momentum fraction, x, in data for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$. At $x \gg 1$, it can be approximated well by a power law. Its distribution is used to model single-misidentified backgrounds as in eq. (4.6).



can be made with a spline-like third-order polynomial, which also has an analytical solution for eq. (4.6). This function is given fully by being fixed to zero at $x_{\min}$, extends to $x_{\min} + \Delta x$ and that it is continuous both for its function and its first derivative.

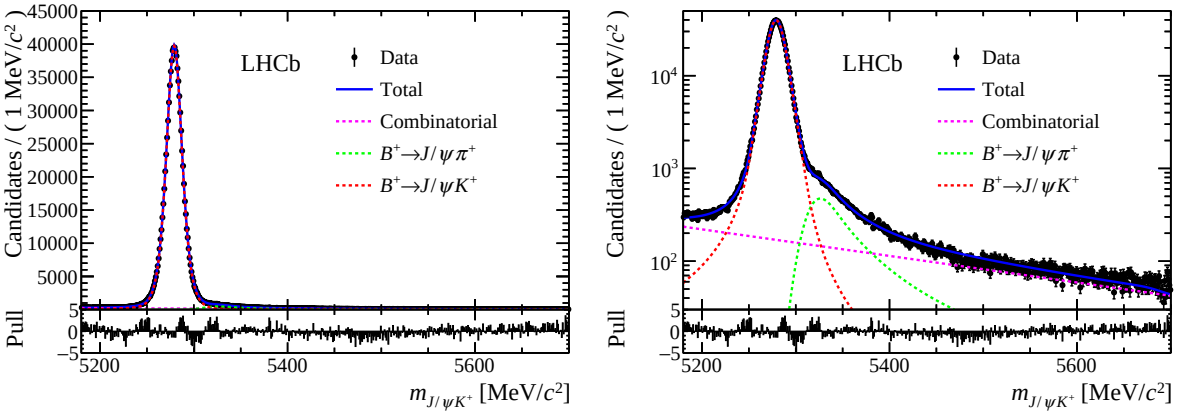
Connecting both the spline and power-law option, an analytical PDF is constructed, called a misID Gaussian, for the distribution of $m_{J/\psi K^\pm}$ for $B^+ \to J/\psi(\to \mu^+\mu^-)\pi^+$. It is fully defined in appendix A.6. The PDF depends on the mean, μ, and width, σ, of the invariant-mass resolution under the correct hypothesis, the masses of the misidentified particle and the hypothesis and the properties of the distribution of x: $x_{\min}$, $x_{\max}$, Δx and the power of the power law. The requirements on the momenta of the particles give $x_{\min}$ and $x_{\max}$ and are fixed accordingly. The range of the spline, Δx, is fixed from simulation of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$. The power is left free in the fit. The masses of the pion and kaon from the PDG fix the transformation of eq. (4.4) [59]. The parameters μ and σ are shared with the core Gaussian of the Hypatia function of the signal mode.

The full PDF describing the $m_{J/\psi K^\pm}$ distribution is given by a sum of the Hypatia function for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, an exponential for the combinatorial background and a misID Gaussian for the $B^+ \to J/\psi(\to \mu^+\mu^-)\pi^+$ background, all multiplied with their respective relative yields. The maximisation of the extended likelihood, based on the PDFs of its components as described above, is performed on $m_{J/\psi K^\pm}$ in the window of $m_{J/\psi K^\pm} \in (5180, 5680)\,\mathrm{MeV}/c^2$. Accordingly, both the distribution as well as the resulting PDF after likelihood maximisation are shown in fig. 4.9. The resulting yield of $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is,

$$913074 \pm 1106\ . \tag{4.7}$$

This includes the systematic uncertainties from the description of the PDFs, as all major uncertainties are included with Gaussian priors in the likelihood. This regards mostly the parameters of the tails of the signal peak, which are the only parameters with significant correlation with the background shapes. While the pulls of the fit with respect to the data shows clearly a non-Gaussian structure, it does not result in large systematic uncertainties, as these come solely from the description of the signal distribution. Since the signal over background ratio is very low, any relative uncertainty on the background gets diluted by the fact that the total amount is already low. In any case, both statistical and systematic uncertainties are negligible with respect to the uncertainties on the efficiency

Figure 4.9: Distribution of $m_{J/\psi K^+}$ for the normalisation mode $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ with fit overlaid: normal scale (left) and log scale (right). Pulls of the fits with respect to the data is shown in the bottom.



determination.

4.3 $B^0 \to K^+\pi^-$

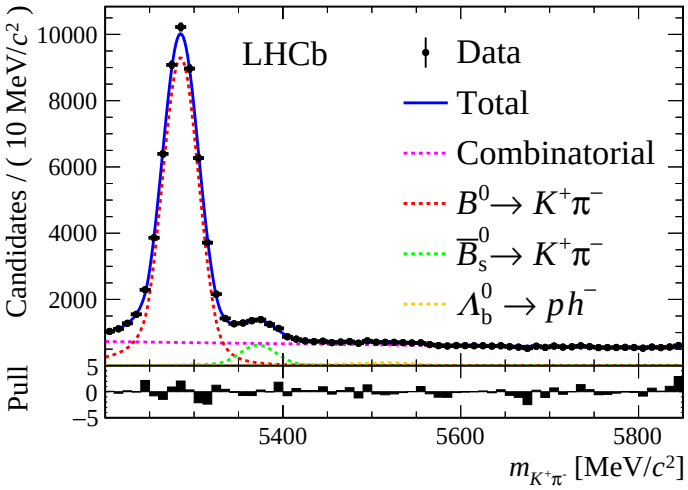
The yield of the normalisation channel $B^0 \to K^+\pi^-$ is determined from the distribution of the invariant mass of its final-state particles, in this case the $K^+\pi^-$ combination.

To describe the PDF of the $B^0 \to K^+\pi^-$ component, a strategy similar to $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is used. As both the branching fraction and the efficiency of reconstructing and selecting $B^0 \to K^+\pi^-$ with respect to $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is lower, the statistics will be quite lower. Therefore, the description of the invariant mass with a double-sided Crystal Ball is sufficient. The mean and width of the Gaussian core are left free in the fit and the remaining parameters are constrained with Gaussian priors in the likelihood. The values of the Gaussian priors are taken from a likelihood fit to simulation of $B^0 \to K^+\pi^-$.

Like $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, the main type of background is combinatorial. This is described by an exponential distribution of which the parameter is free in the fit. Regarding backgrounds of which the particles come from one b-hadron decay, the $m_{K^+\pi^-}$ spectrum contains a relatively small but non-negligible amount of the decay mode $\overline{B}{}^0_s \to K^+\pi^-$. Both the relative production fraction and the branching fraction is about a factor four smaller. The same function is used to describe its PDF, albeit with a fixed shift in the mean given by the difference of the B^0_s and the B^0 mass from the PDG [59]. Aside from the signal mode and the B^0_s mode, the distribution of $m_{K^+\pi^-}$ contains also a small contribution of a single-misidentified background, $\Lambda^0_b \to p\pi^-$. This contribution is described by a double-sided Crystal Ball function, where its parameters are constrained to Gaussian priors in the likelihood to the values obtained to a fit to simulation of $\Lambda^0_b \to p\pi^-$. Other two-body charged hadron decays from b-hadrons are found to be negligible due to the particle-identification requirements.

The resulting full PDF describing the distribution of $m_{K^+\pi^-}$ in the window of $m_{K^+\pi^-} \in (5200, 5850)\,\text{MeV}/c^2$ is given by the sum of the PDF of the components: the decay of $B^0 \to K^+\pi^-$, $\overline{B}{}^0_s \to K^+\pi^-$ and $\Lambda^0_b \to p\pi^-$ and combinatorial backgrounds. It is subsequently used to perform an extended maximum-likelihood fit to obtain the yields of its components.

Figure 4.10: Distribution of $m_{K^+\pi^-}$ for the normalisation mode $B^0 \to K^+\pi^-$ with the PDF after likelihood maximisation overlaid. At the bottom of the figure, the pulls of the PDF after the fit with respect to the data are shown.



The distribution of $m_{K^+\pi^-}$ and the PDF after likelihood maximisation are shown in fig. 4.10. The resulting yield of $B^0 \to K^+\pi^-$ is,

$$49907 \pm 277 \ . \tag{4.8}$$

This includes systematic uncertainties on the description of invariant mass of the signal mode by the use of Gaussian priors. Both statistical and systematic uncertainties are negligible with respect to the efficiency determination.

4.4 Normalisation factors and normalisation-ratio cross-check

Beside the normalisation factors described hitherto, two more effects need to be taken into account. First, the efficiencies for signal, including $B^0 \to e^\pm\mu^\mp$, are determined with simulation of $B^0_s \to e^\pm\mu^\mp$. The difference with respect to $B^0 \to e^\pm\mu^\mp$ is mainly the invariant mass. This results in about a 3% lower general selection efficiency for $B^0 \to e^\pm\mu^\mp$. Differences in kinematics are negligible [60]. Second, errors during data processing result in a different luminosity between channels. These are of the order of a percent and are included in the final normalisation factors. The uncertainty on this factor is negligible.

With all factors going into the normalisation obtained, summarised in table 4.5, eq. (4.1) can be used to obtain the branching fractions of $B^0 \to e^\pm\mu^\mp$ and $B^0_s \to e^\pm\mu^\mp$ from their respective yields. These result in normalisation constants, *i.e.* the factors between the yield and the branching fractions as given in eq. (4.1), of $(2.48 \pm 0.17) \cdot 10^{-10}$ and $(6.16 \pm 0.23) \cdot 10^{-11}$ for B^0_s and B^0 respectively. These go directly into the likelihood fit for the signal modes, discussed in chapter 5.

As a cross-check to the normalisation procedure, the ratio of branching fractions of

Table 4.5: Summary of factors going into the normalisation eq. (4.1). Branching fractions are from the PDG [59]. The relative production fraction is from [60]. An additional factor comes from a difference in data processing, effectively giving a luminosity difference for $B^0 \to K^+\pi^-$ of 1.3%.

	$B_s^0 \to e^\pm\mu^\mp$	$B^0 \to e^\pm\mu^\mp$	$B^+ \to J/\psi(\to \mu^+\mu^-)K^+$	$B^0 \to K^+\pi^-$
$\mathcal{B}$	-	-	$(6.10 \pm 0.20) \cdot 10^{-5}$	$(1.96 \pm 0.20) \cdot 10^{-5}$
ϵ	$(2.29 \pm 0.05)\%$	$(2.22 \pm 0.05)\%$	$(2.215 \pm 0.035)\%$	$(0.360 \pm 0.021)\%$
f_q/f_d	$(25.9 \pm 1.5)\%$	1	1	1
N	-	-	913074 ± 1106	49907 ± 277

$B^0 \to K^+\pi^-$ to $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ is determined. The measured value is,

$$R_{\text{norm}} = \frac{N_{B^0\to K^+\pi^-} \cdot \varepsilon_{B^+\to J/\psi(\to\mu^+\mu^-)K^+}}{N_{B^+\to J/\psi(\to\mu^+\mu^-)K^+} \cdot \varepsilon_{B^0\to K^+\pi^-}} = 0.332 \pm 0.002\,(\text{stat}) \pm 0.020\,(\text{syst})\ , \qquad (4.9)$$

where the statistical uncertainty is essentially the uncertainty of the yields of the normalisation channels; the systematic uncertainty is mostly the uncertainty from the efficiency determinations. Included in the ratio calculation is the difference in data processing of 1.3%. The value of R_{norm} is compatible with the value of the PDG: 0.321 ± 0.013 [59].

Chapter 5

Likelihood fit

To extract the branching fractions of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ from the dataset, a PDF describing the dataset has to be constructed, with the branching fractions as parameters. The goals is to maximise the likelihood of this PDF such that it gives the best description of the dataset and therefore the best estimate of the branching fractions.

The PDF will be a function of the three observables. First, the observable that is most suitable is the invariant mass of the $e^\pm\mu^\mp$ combination, $m_{e^\pm\mu^\mp}$, since the signal peaks in its distribution and the background does not. In addition, the distributions of the individual signal and background components can be modelled well. Second, as the other non-particle-identification observables that have significant distinguishing power are combined into the BDT classifier, the invariant-mass distributions are categorised in bins of the BDT classifier to exploit its separation power maximally. In addition, a split has been made in the bremsstrahlung category, `HasBremAdded`, as particles that are not real electrons are much less likely to have a bremsstrahlung-photon candidate assigned. The ranges of these observables are given by $m_{e^\pm\mu^\mp} \in [4900, 5850]\,\text{MeV}/c^2$, BDT $\in [0.25, 1.0]$ and `HasBremAdded` $\in [0, 1]$.

The extended likelihood constructed with the PDF is given by,

$$\mathcal{L} = \frac{(n_{B^0} + n_{B_s^0} + n_b)^N e^{-(n_{B^0}+n_{B_s^0}+n_b)}}{N!} \prod^{N} f(\vec{x}|\lambda) \prod^{|\subset\lambda|} g(\subset\lambda|\theta) \ , \tag{5.1}$$

where $\vec{x} = (m_{e^\pm\mu^\mp}, \text{BDT}, \texttt{HasBremAdded})$ is an entry in the signal dataset of size N; f is the product of the PDFs in $m_{e^\pm\mu^\mp}$, BDT and `HasBremAdded` of $B_s^0 \to e^\pm\mu^\mp$, $B^0 \to e^\pm\mu^\mp$ and the backgrounds; the set of parameters of the PDFs are denoted by λ; the yields n_{B^0} and $n_{B_s^0}$ depend on their corresponding branching fraction with eq. (4.1) and the parameter n_b is the sum of the yields of the backgrounds [59, 65]. The determination of the PDFs for $B_s^0 \to e^\pm\mu^\mp$ and $B^0 \to e^\pm\mu^\mp$ is described in section 5.1. The composition, yields and distributions of the backgrounds will be discussed in section 5.2. Some parameters of the PDF are not left free in the fit (*e.g.* some yields of the backgrounds). They are represented by θ in eq. (5.1). To incorporate the statistical and systematic uncertainties from their calibration, the Gaussian PDF $g(\subset\lambda|\theta)$ is added as a prior to the likelihood [67]. These parameters are therefore not fixed, but minimised in the likelihood fit while constraining them with their uncertainty.

The results of the likelihood maximisation are discussed and shown in section 5.3. One candidate in the dataset that stands out in the fit is highlighted with the LHCb

event display and further discussed in section 5.4 to illustrate the connection between the detector and its raw data on one side and analysis on the other.

5.1 Signal

For signal (both for $B_s^0 \to e^\pm\mu^\mp$ and $B^0 \to e^\pm\mu^\mp$), the PDF is described by,

$$f^{\text{sig}}(m_{e^\pm\mu^\mp}, k, l) = f_k^{\text{brem}} f_{k,l}^{\text{BDT}} f_k^{\text{mass}}(m_{e^\pm\mu^\mp}) \ , \tag{5.2}$$

where k and l are the indices of the bremsstrahlung category and of the BDT classifier bins respectively. The parameters f_k^{brem} and $f_{k,l}^{\text{BDT}}$ are the fraction of the total yield in bin k and k, l respectively, essentially describing the binned PDF in these variables. The term $f_k^{\text{mass}}(m_{e^\pm\mu^\mp})$ is the unbinned PDF of the $m_{e^\pm\mu^\mp}$ distribution per bremsstrahlung category. The determination of $f_k^{\text{mass}}(m_{e^\pm\mu^\mp})$, $f_{k,l}^{\text{BDT}}$ and f_k^{brem} is discussed in section 5.1.1, section 5.1.2 and section 5.1.3, respectively.

5.1.1 Invariant mass

The invariant-mass PDF for $B_s^0 \to e^\pm\mu^\mp$ is determined with simulation, but includes a small correction to the resolution from a data-driven method. For $B^0 \to e^\pm\mu^\mp$, the simulation of $B_s^0 \to e^\pm\mu^\mp$ is used as a proxy, where the PDF is corrected by a shift of $m_{B^0} - m_{B_s^0} = -(87.24 \pm 0.21)\,\text{MeV}/c^2$ [59]. The analytical function used for the description of the PDF is a double-sided Crystal Ball, defined in eq. (A.1). Essentially, it is a Gaussian distribution at its core with power-law tails on both sides. The latter allow for the incorporation of the tails of the PDF due to bremsstrahlung losses and its recovery. The invariant-mass distributions in simulation of $B_s^0 \to e^\pm\mu^\mp$ are shown, separately for bremsstrahlung categories, with the signal PDF resulting from the likelihood fit overlaid, in fig. 5.1.

The data-driven method to correct the resolution is performed by comparing momentum resolutions of calibration channels, both for muons and electrons, in data and simulation and combining these to form a correction for $m_{e^\pm\mu^\mp}$ from simulation to data. The derivation of the correction starts with the assumption that the final state particles are highly relativistic, *i.e.* $E \approx p$. This way, the invariant mass of $e^\pm\mu^\mp$, $m_{e^\pm\mu^\mp}$, and in general any two highly-relativistic particle combinations, is given by,

$$m_{e^\pm\mu^\mp} = \sqrt{(p_{e^\pm}^\nu + p_{\mu^\mp}^\nu)^2} \simeq \sqrt{2p_e p_\mu(1 - \cos\theta)} \ , \tag{5.3}$$

where $p_{e^\pm}^\nu$ and $p_{\mu^\mp}^\nu$ are the four-momenta of the electron and the muon respectively. The variables p_e and p_μ are the absolute momenta of the electron and the muon respectively and θ is the opening angle between the muon and the electron.

Using eq. (5.3) and neglecting the opening-angle resolution, the invariant-mass resolution is given by,

$$\left(\frac{\sigma_m}{m}\right) = \frac{1}{2}\sqrt{\left(\frac{\sigma_{p_e}}{p_e}\right)^2 + \left(\frac{\sigma_{p_\mu}}{p_\mu}\right)^2} \ , \tag{5.4}$$

where σ_m, σ_{p_e} and σ_{p_μ} are the resolutions of the invariant mass, electron absolute momentum and the muon absolute momentum respectively. By adapting eq. (5.4) and subsequently applying it on the invariant mass of the decay of $J/\psi \to e^+e^-$ and $J/\psi \to \mu^+\mu^-$, the momentum resolutions of the electron and muon can be extracted. These momentum resolutions can be used to calculate the invariant-mass resolution of the $e^\pm\mu^\mp$ combination directly from the invariant-mass resolutions of $J/\psi \to e^+e^-$ and $J/\psi \to \mu^+\mu^-$. This can be done for both data and simulation in the calibration channel, leading to the correction factor,

$$C^{m_{e^\pm\mu^\mp}}_{\text{data/sim}} = \frac{\left(\frac{\sigma_m}{m}\right)_{e^\pm\mu^\mp;\ \text{data}}}{\left(\frac{\sigma_m}{m}\right)_{e^\pm\mu^\mp;\ \text{sim}}} = \frac{\sqrt{\left(\frac{\sigma_m}{m}\right)^2_{e^+e^-;\ \text{data}} + \left(\frac{\sigma_m}{m}\right)^2_{\mu^+\mu^-;\ \text{data}}}}{\sqrt{\left(\frac{\sigma_m}{m}\right)^2_{e^+e^-;\ \text{sim}} + \left(\frac{\sigma_m}{m}\right)^2_{\mu^+\mu^-;\ \text{sim}}}} \ , \tag{5.5}$$

where the superscript denotes if it is the invariant-mass resolution in data or simulation (sim) and the subscript denotes which two-particle combination it is: $e^\pm\mu^\mp$ for $B^0_s \to e^\pm\mu^\mp$ and $B^0 \to e^\pm\mu^\mp$, $\mu^+\mu^-$ for $J/\psi \to \mu^+\mu^-$ and e^+e^- for $J/\psi \to e^+e^-$.

The decays of $B^+ \to J/\psi(\to e^+e^-)K^+$ and $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ are used as calibration channels. The invariant-mass resolutions are extracted by performing a likelihood fit on the e^+e^- or $\mu^+\mu^-$ invariant-mass on data and simulation. A double-sided Crystal Ball is used as a PDF describing the signal, $J/\psi \to e^+e^-$ or $J/\psi \to \mu^+\mu^-$, and an exponential describing the combinatorial background. The width of the Gaussian cores of the double-sided Crystal Balls are used for the invariant-mass resolutions. The fits result in a correction factor, as defined in eq. (5.5), of,

$$C^{m_{e^\pm\mu^\mp}}_{\text{data/sim}} = 1.09 \pm 0.03 \ , \tag{5.6}$$

which is used to multiply the values of the widths of the Gaussian cores in the signal PDFs as shown in fig. 5.1. Subsequently, the corrected PDFs are used, both for $B^0_s \to e^\pm\mu^\mp$ as for $B^0 \to e^\pm\mu^\mp$, to define $f^{\text{mass}}_k(m_{e^\pm\mu^\mp})$ as given in eq. (5.2).

To account for systematic effects, the statistical uncertainties on the parameters of the PDF from the fit to simulation and the correction factor from eq. (5.6) are used as Gaussian constraints in eq. (5.1).

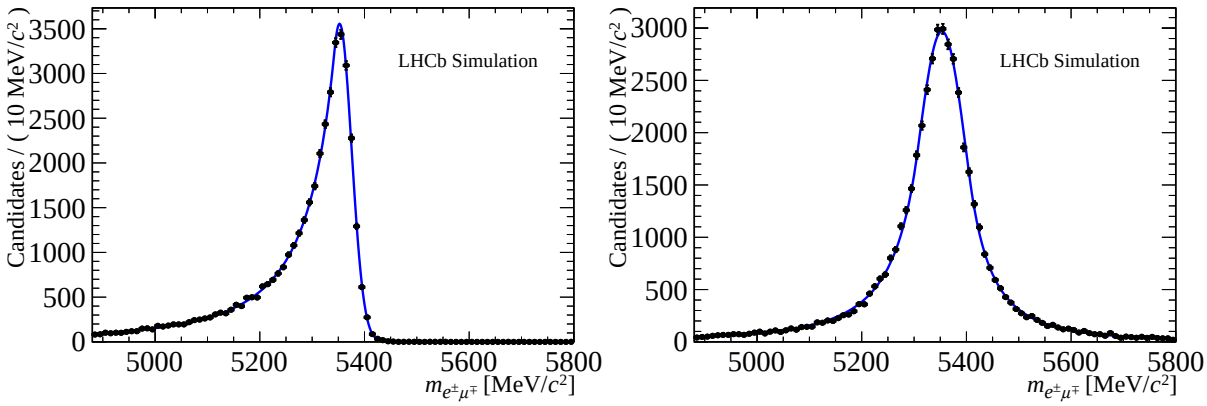
5.1.2 BDT

For the determination of the PDF of the BDT classifier, *i.e.* the factors $f^{\text{BDT}}_{k,l}$ from eq. (5.2), the choice of input variables opens up the possibility to calibrate $f^{\text{BDT}}_{k,l}$ on data on a decay channel that is a close proxy. Since there are no particle-identification variables used in the BDT, the decay of $B^0 \to K^+\pi^-$ is a good proxy. In essence, the relative yields of $B^0 \to K^+\pi^-$ in BDT classifier bins, corrected with any residual differences between $B^0_{(d/s)} \to e^\pm\mu^\mp$ and $B^0 \to K^+\pi^-$ (obtained from simulation), gives the PDF of the BDT classifier. The likelihood fit to $m_{K^+\pi^-}$ is the same as described in section 4.3 and therefore will not be discussed further.

The focus will be on the main question of the evaluation of the differences between $B^0_{(d/s)} \to e^\pm\mu^\mp$ and $B^0 \to K^+\pi^-$: the particle-identification selection, the trigger selection and kinematics related to bremsstrahlung.

For particle identification, it is chosen to have a `DLLK` $> \kappa$ for the kaon candidate and `DLLK` $< -\kappa$ for the pion candidate, where $\kappa = 5$ is chosen as nominal value. The

Figure 5.1: Distributions of the invariant mass of $e^{\pm}\mu^{\mp}$ from simulated $B_s^0 \to e^{\pm}\mu^{\mp}$ decays, $m_{e^{\pm}\mu^{\mp}}$, for bremsstrahlung categories: `!HasBremAdded` (left) and `HasBremAdded` (right). The signal PDF resulting from the likelihood fit is overlaid.



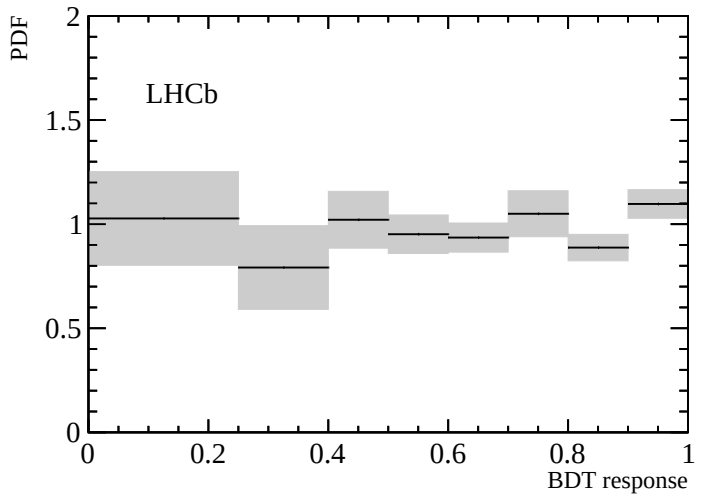
purpose of this selection is to get rid of backgrounds like $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$. The efficiency of this requirement, in BDT bins, is determined with `PIDCalib` on the $D^{*+} \to D^0(\to K^+\pi^-)\pi^+$ calibration sample, where the efficiencies are calculated in bins of p and η. The efficiencies have been recalculated in different binning schemes to determine systematic uncertainties. As a cross-check, the efficiency-corrected yields per BDT bin are redetermined in a range of $\kappa \in [5, 10]$. The result is consistent along the whole range and therefore no systematic uncertainty is assigned for this.

In case of the trigger selection, the most efficient choice at `L0` level would be `L0Hadron`, but since the transverse energy-deposit requirements are a lot harsher than the requirements on `L0Muon` and `L0Electron` for $B^0_{(d/s)} \to e^{\pm}\mu^{\mp}$ candidates, this would introduce large corrections for trigger differences. For this reason, a `TIS` selection is chosen, which does not have such large differences. To avoid biases from overlap between `TIS` and `TOS`, `TIS` from `L0Hadron` is not used. At `Hlt1` level, a collection of lines is used for a `TIS` selection as well. Since at `Hlt2` level the trigger selection can be similar to $B^0_{(d/s)} \to e^{\pm}\mu^{\mp}$, a `TOS` selection is used. The differences in trigger efficiency between the BDT calibration with $B^0 \to K^+\pi^-$ and $B_s^0 \to e^{\pm}\mu^{\mp}$ have been determined with simulation for each bin in BDT classifier. No dependence on the BDT classifier is seen and the statistical uncertainty from this determination is taken as systematic uncertainty on $f_{k,l}^{\mathrm{BDT}}$.

As the kinematics per bremsstrahlung category is different, the factors $f_{k,l}^{\mathrm{BDT}}$ are determined per bin in bremsstrahlung category in simulation of $B_s^0 \to e^{\pm}\mu^{\mp}$.

To summarise, the BDT PDF factors $f_{k,l}^{\mathrm{BDT}}$ are determined using yields extracted with likelihood fits to $B^0 \to K^+\pi^-$ data in bins of the BDT classifier and bremsstrahlung category. Subsequently, the yields are corrected for the differences in particle-identification and trigger selection. The statistical and systematic uncertainties from the determination are included in eq. (5.1) as Gaussian constraints. The resulting PDF for the BDT for the `HasBremAdded` category is shown in fig. 5.2. The resulting distribution is consistent with being flat, within the combination of statistical and systematic uncertainties.

Figure 5.2: Distribution of the BDT classifier for $B^0_s \to e^\pm\mu^\mp$ in the category where bremsstrahlung is recovered, determined with fits to the $B^0 \to K^+\pi^-$ control mode, corrected with differences between $B^0 \to K^+\pi^-$ and $B^0_s \to e^\pm\mu^\mp$ with data-driven efficiencies for particle identification and other differences with simulation. Black lines denote central values, grey bands denote the combination of statistical and systematic uncertainties. Each bin is normalised to its width, therefore the factors $f^{\rm BDT}_{k,l}$ are obtained by multiplying with the bin width.



5.1.3 Bremsstrahlung category

The likelihood is also binned in bremsstrahlung category, `HasBremAdded`, and thus how much signal and background ends up in which category needs to be determined. For other analyses at LHCb, it is known that this fraction of electrons that have bremsstrahlung added, $f_{\rm brem}$, is well reproduced in simulation. For $B^0_s \to e^\pm\mu^\mp$, the fraction determined with simulation is,

$$f_{\rm brem} = (57.96 \pm 0.17\,({\rm stat}) \pm 0.20\,({\rm syst}))\% \ , \tag{5.7}$$

where the statistical uncertainty comes from the simulation sample size and the systematic uncertainty is from data and simulation differences determined in [18]. Therefore, $f^{\rm brem}_k = f_{\rm brem}$ and $f^{\rm brem}_k = 1 - f_{\rm brem}$ for k corresponding to `HasBremAdded` and `!HasBremAdded` respectively.

As a cross-check, $f_{\rm brem}$ is evaluated for $B^+ \to J/\psi(\to e^+e^-)K^+$ decays in data and compared to its corresponding simulation sample. For simulation, the determination is done in two ways: one with default simulation and one where the occupancy of the events has been reweighed to match the data sample, as not-signal related bremsstrahlung-photon candidates are more likely in high occupancy events. The resulting values of $f_{\rm brem}$, determined with the relative yields of the fit to the bremsstrahlung categories, are listed in table 5.1, together with $f_{\rm brem}$ from simulation. It does not show large differences and agrees with the determination from [18].

Table 5.1: Fraction of electrons that have bremsstrahlung recovered, f_{brem}, in data and simulation, both default and reweighed in occupancy, *i.e.* `nSPDHits`, for $B^+ \to J/\psi(\to e^+e^-)K^+$. The uncertainties for the two simulation determinations are shared, as they are dominated by the reweighing procedure. Therefore the results do not show significant differences, *i.e.* not larger than 2σ.

Data	Simulation	Reweighed simulation
$(47.94 \pm 0.17)\%$	$(47.1 \pm 0.4)\%$	$(47.8 \pm 0.4)\%$

5.2 Backgrounds

Having determined the PDF of $m_{e^\pm\mu^\mp}$, the BDT classifier and `HasBremAdded` for $B^0_s \to e^\pm\mu^\mp$ and $B^0_s \to e^\pm\mu^\mp$, what remains is to determine the contribution of backgrounds on the distribution of these observables.

In the case of combinatorial background, the invariant-mass PDF is described by an exponential function. The reason why combinatorial backgrounds behave exponentially in terms of their invariant mass, is that the distribution of the transverse momentum of particles drops exponentially towards higher p_{T}, while its pseudo-rapidity distribution is fairly flat. The resulting invariant mass of random, *i.e.* uncorrelated, combinations of these particles is exponential. Since a priori it is not known how the shapes depend on the BDT classifier and the bremsstrahlung category, it was chosen to use independent parameters of the PDF in each BDT bin and bremsstrahlung category. The parameters and the yields are free in the fit.

For the remaining possible backgrounds, different assumptions are possible, since the $e^\pm\mu^\mp$ combination may not be random and come from one b-hadron decay. These backgrounds can be categorised into three types.

Types of b-hadron backgrounds

- **Misidentification** A b-hadron decay to two charged particles where at least one is misidentified. An example of such a decay is $B^0 \to K^+\pi^-$.
- **Partially reconstructed** A b-hadron decay where there is a muon and an electron in the final state, together with additional final-state particles. For example, where two final-state particles are not reconstructed, *e.g.* a neutrino and a positron in the case of the decay $B_c^+ \to J/\psi\ (e^+e^-)\mu^+\nu_\mu$.
- **Combination** A b-hadron decay with a combination of at least one misidentified particle with at least one final-state particle that is not reconstructed. For example $B^0 \to \pi^-\mu^+\nu_\mu$, where the neutrino is not reconstructed and the π^- is misidentified as an electron.

What has to be checked first is which b-hadron decays have a significant yield within the signal selection. To calculate the yield of the background, a normalisation method is applied similar to eq. (4.1). The $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ normalisation channel is used as sole normalisation channel, since this channel has the most similarities with the significant backgrounds, which, as will be shown, have a muon in the final state. The difference

with the procedure in eq. (4.1) is that not a branching fraction is calculated, but the expected yield, which can be done since the branching fractions of the significant b-hadron background modes are known. The yield of background X is given by,

$$N_X = N_{B^+\to J/\psi K^+} \frac{f_q}{f_u} \frac{\mathcal{B}(X)}{\mathcal{B}(B^+ \to J/\psi K^+)} \frac{\epsilon(X)}{\epsilon(B^+ \to J/\psi K^+)} \,, \tag{5.8}$$

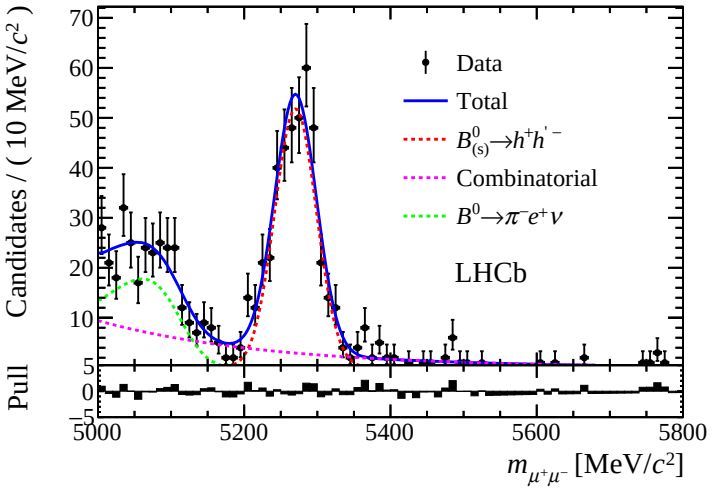
where $N_{B^+\to J/\psi K^+}$, $\frac{f_q}{f_u}$, $\mathcal{B}(B^+ \to J/\psi K^+)$ and $\epsilon(B^+ \to J/\psi K^+)$ are the yield, relative production fraction, branching fraction and efficiency of the normalisation channel $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ respectively, as reported in eq. (4.1). The branching fraction of the background, $\mathcal{B}(X)$, is taken from the PDG [59]. The efficiency of reconstructing and selecting the background, $\epsilon(X)$, is calculated using simulation, except for the particle-identification efficiency, which is done using the data-driven `PIDCalib` method, as discussed in section 4.1.3.

The decays that turned out to be relevant, *i.e.* have more than one expected event with its uncertainty as margin, are $\Lambda_b^0 \to p\mu^-\nu$ and $B^0 \to \pi^-\mu^+\nu$. An expected yield of 82 ± 39 and 55 ± 3 has been determined for the full BDT range respectively.

As discussed in section 3.2.3, the $B \to h^+h^-$ backgrounds were potentially the most dangerous, as it peaks very similar to $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ in the invariant mass, $m_{e^\pm\mu^\mp}$, and behaves practically the same in the BDT. While from the background normalisation procedure no relevant amount, 0.11 ± 0.02, is expected for the total of all $B \to h^+h^-$ decays, due to its dangerous nature, a cross-check has been performed. As by far the largest difference between signal and $B \to h^+h^-$ backgrounds is the particle identification, the efficiencies of the particle-identification selection are cross-checked. With the nominal method, the double-misidentification efficiency is determined fully with `PIDCalib`. For the purpose of the cross-check, the signal dataset is adapted by only setting the electron-identification requirement and omitting the muon identification and replacing it with a pion identification requirement. This results in a dataset with a sufficient number of $B \to h^+h^-$ decays. In turn this can be compared to the expected number of $B \to h^+h^-$ backgrounds in the signal sample by only needing to determine the efficiency of the muon identification requirement and light pion identification requirement. One could reverse the procedure to also cross-check the muon identification efficiency, but this is not needed as it is already done for the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction [54]. The invariant mass of the two charged particles assuming the muon hypothesis is shown for the cross-check in fig. 5.3. With a extended likelihood fit the yield of the $B \to h^+h^-$ backgrounds can be determined and the PDF resulting from the fit is overlaid. The yield corrected with the particle-identification efficiencies determined with `PIDCalib` for the new selection is compatible with the expected yield of the nominal method, which resulted in a non-significant amount of $B \to h^+h^-$ backgrounds.

For the significant backgrounds, $\Lambda_b^0 \to p\mu^-\nu$ and $B^0 \to \pi^-\mu^+\nu$, the PDFs for $m_{e^\pm\mu^\mp}$, BDT and `HasBremAdded` are determined from simulation by binning in BDT and `HasBremAdded` and using kernel density estimation for $m_{e^\pm\mu^\mp}$ [71]. Essentially, it describes the PDF by a sum of Gaussians, where the mean of each Gaussian is given by a data point in a reference dataset, this case simulation. The expected yields of these backgrounds and their uncertainties are used to constrain their yields in the likelihood fit. Systematic effects from the determination of the shape of the exclusive backgrounds is negligible, since the expected yields are small while their relative uncertainty is fairly large and in addition, the distributions are not strongly peaking, but have a more exponential

Figure 5.3: Distribution of the invariant mass of the two charged particles under the muon hypothesis, $m_{\mu^+\mu^-}$, for the dataset of the cross-check of the $B \to h^+h^-$ backgrounds. The muon selection is omitted and changed in a light pion-identification selection. The yield of the $B \to h^+h^-$ backgrounds, corrected for the muon-identification efficiency in the $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ sample and the pion-identification selection, is compatible with the nominal method. No significant amount of $B \to h^+h^-$ is expected. The pulls of PDF resulting from fit with respect to the data are shown at the bottom.



distributed behaviour. This dilutes any mismodelling in the resolution. Note that these significant backgrounds all contain an electron misidentification. Therefore, the splitting in `HasBremAdded` is a cross-check in itself, as a lot less of these backgrounds is expected in the `HasBremAdded` category.

5.3 Fit results

Using the normalisation procedure in chapter 4 and the PDFs of the signal and backgrounds, discussed in the previous sections, the branching fractions can be extracted. The fits are performed simultaneously for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$. For $B_s^0 \to e^\pm\mu^\mp$, the two extreme cases of the effective lifetime have been determined separately.

The extended likelihood from eq. (5.1) is maximised by minimising $-\log(\mathcal{L})$ with the MINUIT package, using MIGRAD for the minimization and HESSE for the determination of the uncertainties [72, 73]. The dataset contains 476 candidates and the branching fractions resulting from the fit are reported in table 5.2. The distributions of $m_{e^\pm\mu^\mp}$ in bins of the BDT classifier and `HasBremAdded` are shown in fig. 5.5. To highlight the most sensitive region, the distribution of $m_{e^\pm\mu^\mp}$ integrated over the BDT classifier range of 0.7 to 1.0 and `HasBremAdded` is shown in fig. 5.4.

No significant excesses are found. Only the $B_s^0 \to e^\pm\mu^\mp$ decay has a non-significant upward fluctuation due to one candidate in the highest BDT bin with bremsstrahlung added, right at the mass of the B_s^0, *i.e.* around 5367 MeV/c^2. This event will be discussed in section 5.4. What can be noted further is that the goodness-of-fit is sufficient and there are no noticeable differences between the categories with and without bremsstrahlung recovered. This is in a sense a cross-check of the determination of the number of exclusive

Figure 5.4: Distribution of $m_{e^\pm\mu^\mp}$ for BDT classifier larger than 0.7, integrated over `HasBremAdded`. The PDF resulting from the likelihood fit is overlaid. Explanations of the individual components of the PDFs are given in the legend. Due to the result of the fit, the $B^0 \to e^\pm\mu^\mp$ component is not visible.

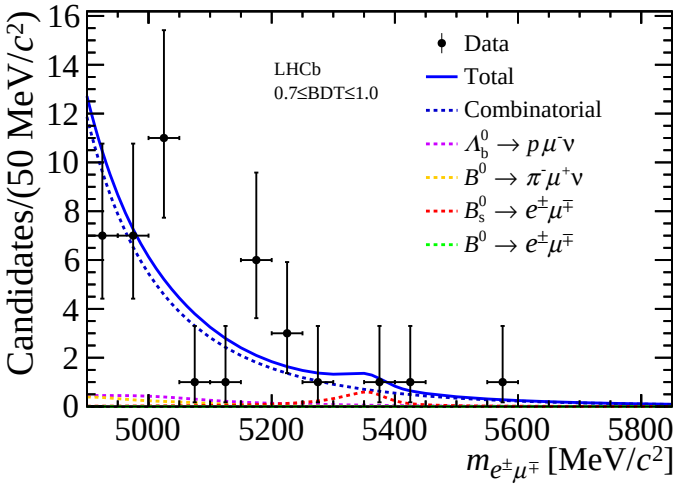


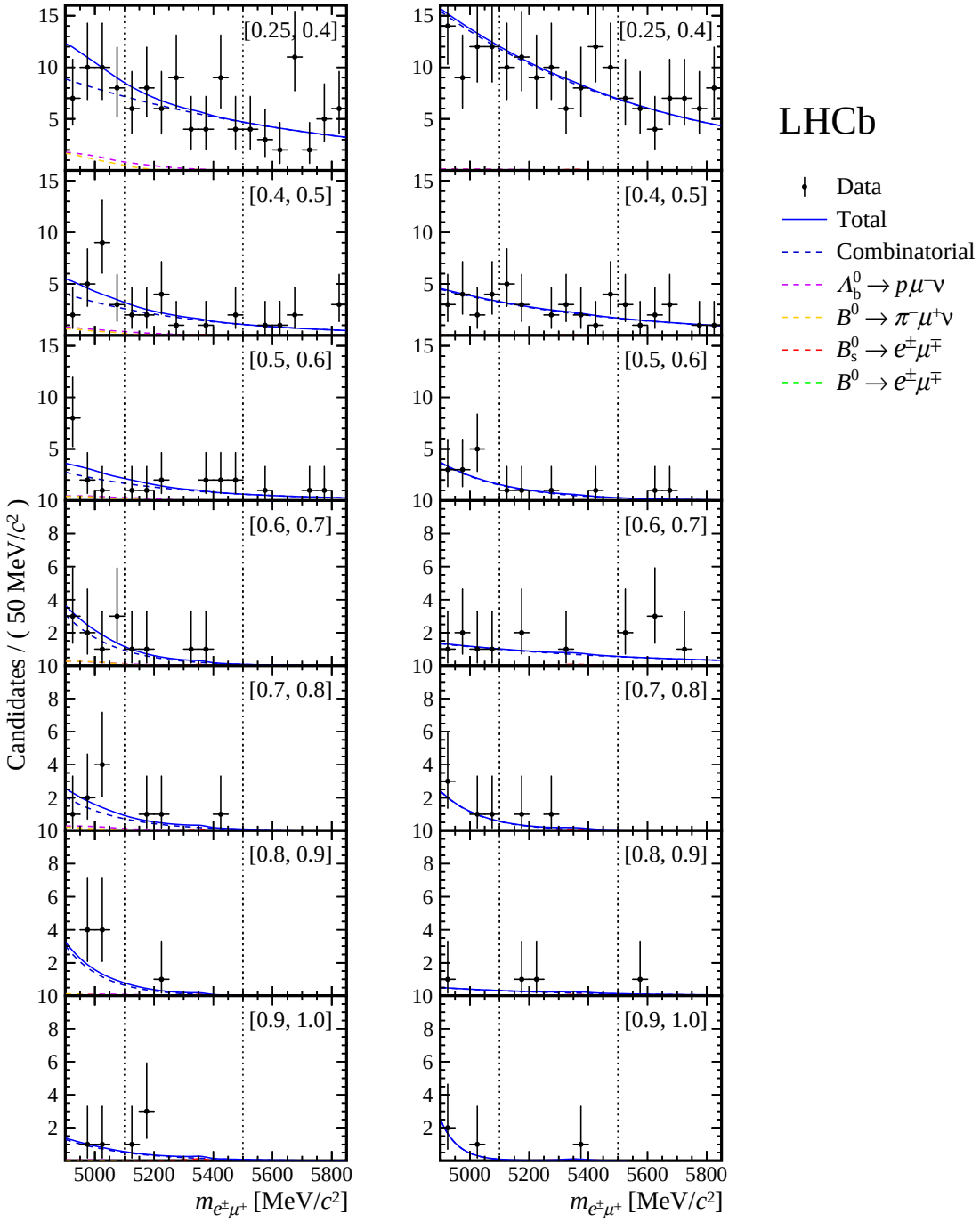
Table 5.2: Branching fractions of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$, both *CP*-even and odd eigenstates, obtained from the extended-likelihood fit shown in fig. 5.5.

Decay channel		Branching fraction
$B_s^0 \to e^\pm\mu^\mp$	*CP*-odd	$(1.3 \pm 1.2) \times 10^{-9}$
	CP-even	$(1.4 \pm 1.3) \times 10^{-9}$
$B^0 \to e^\pm\mu^\mp$		$(0.0 \pm 1.7) \times 10^{-9}$

backgrounds, as they all contain a misidentification of a hadron as an electron, and if there is bremsstrahlung recovered, it is far more likely that the electron candidate is an electron.

In chapter 6, this result will be expressed in terms of an upper limit to the branching fractions.

Figure 5.5: Distribution of $m_{e^\pm\mu^\mp}$ in bins of BDT classifier, denoted by the interval in the top right corners, and `!HasBremAdded` (left) and `HasBremAdded` (right). The PDF resulting from the likelihood fit is overlaid. The individual components of the PDF are explained in the legend. The blinded signal region, *i.e.* the region that was only examined after finalising the selection, is denoted by the gray dashed vertical lines. This region constitutes about 90% of the signal candidates in simulation. Due to the result of the fit, the $B^0 \to e^\pm\mu^\mp$ component is not visible.



5.4 Visualisation: *event-display*

The most striking candidate in the dataset is the one in the highest BDT bin, with bremsstrahlung recovered and close to the value of the B_s^0 mass, clearly visible in fig. 5.5. This event, number 410703310 in run 122547, was recorded in 2012. To show how an event with such a signal-like candidate looks like in the detector, it is shown in the LHCb event display in figs. 5.6 to 5.9.

The color of the tracks are determined according to their most likely hypothesis determined with PID variables, with blue for electrons, red for kaons, orange for pions, purple for protons, and green for muons. Next to the charged tracks, bremsstrahlung photon candidates are shown with blue dashed lines, extrapolated from the start of the associated electron track candidate to the calorimeter. In addition, energy deposits, where the bar scales with the energy, in the electromagnetic calorimeter, in blue, and hadronic calorimeter, in orange, are shown and also green muons station hits are shown.

In figs. 5.6 and 5.8, one can see the muon of the candidate, green line, clearly with the hits in the muon stations. The electron candidate is the blue line with the large cluster in the lower part of the electromagnetic calorimeter. The dashed blue line directly next to it, pointing towards the electromagnetic calorimeter cluster, denotes the bremsstrahlung-photon candidate associated to this electron. In figs. 5.7 and 5.9, one can see the displaced nature of the tracks and the vertex. The PV and secondary (decay) vertex are denoted by the large and small circle respectively. All tracks are extrapolated towards the z position of the PV, so for displaced tracks this is wrong, but does clearly show the size of the impact parameter. The most relevant variables used for the analysis from this candidate are listed in table 5.3. Notable is that both the muon and electron candidate score very high in their particle-identification observables and fired the muon and electron trigger at `L0` respectively. More info on the reconstruction of events, in particular electrons, can be found in chapter 8.

Figure 5.6: Event display of the event with the most signal-like $B_s^0 \to e^{\pm}\mu^{\mp}$ candidate, number 410703310 in run 122547.

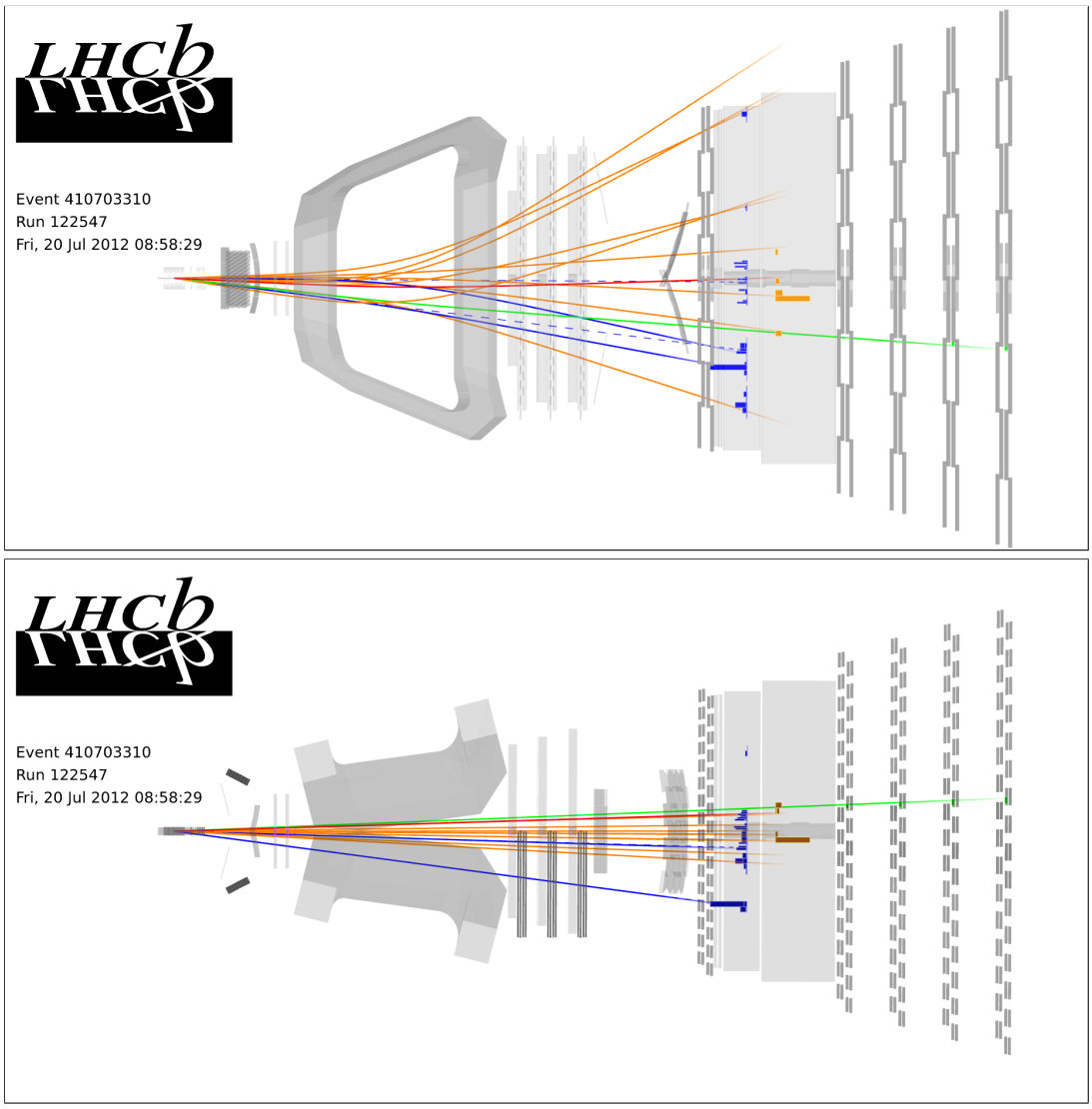


Figure 5.7: Event display, around the VELO, of the event with the most signal-like $B_s^0 \to e^{\pm}\mu^{\mp}$ candidate, number 410703310 in run 122547. In the zoomed pictures, the PV and secondary (decay) vertex are denoted by the large and small circle respectively.

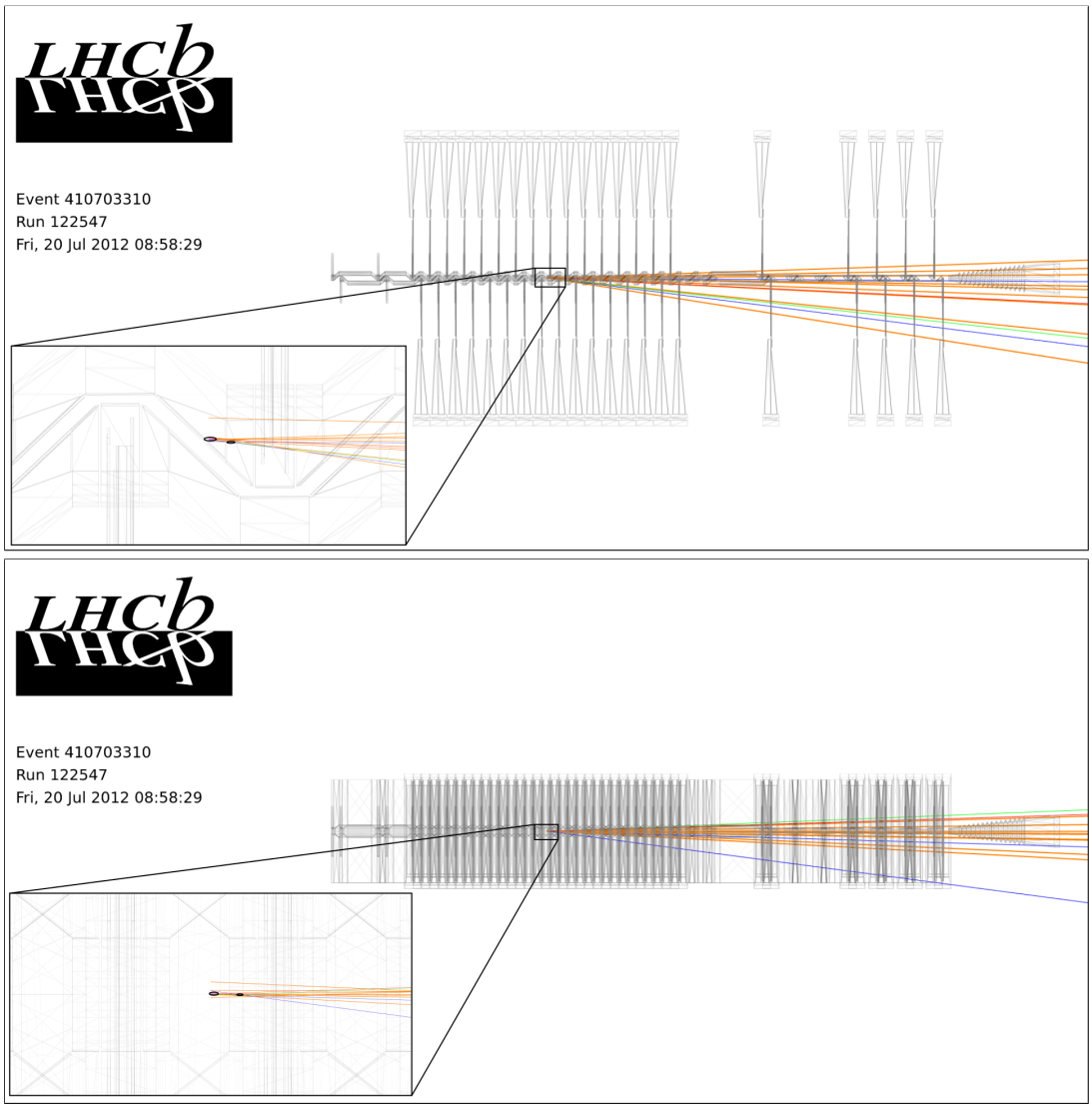


Figure 5.8: Event display, zoomed towards the electromagnetic calorimeter, of the event with the most signal-like $B_s^0 \to e^{\pm}\mu^{\mp}$ candidate, number 410703310 in run 122547.

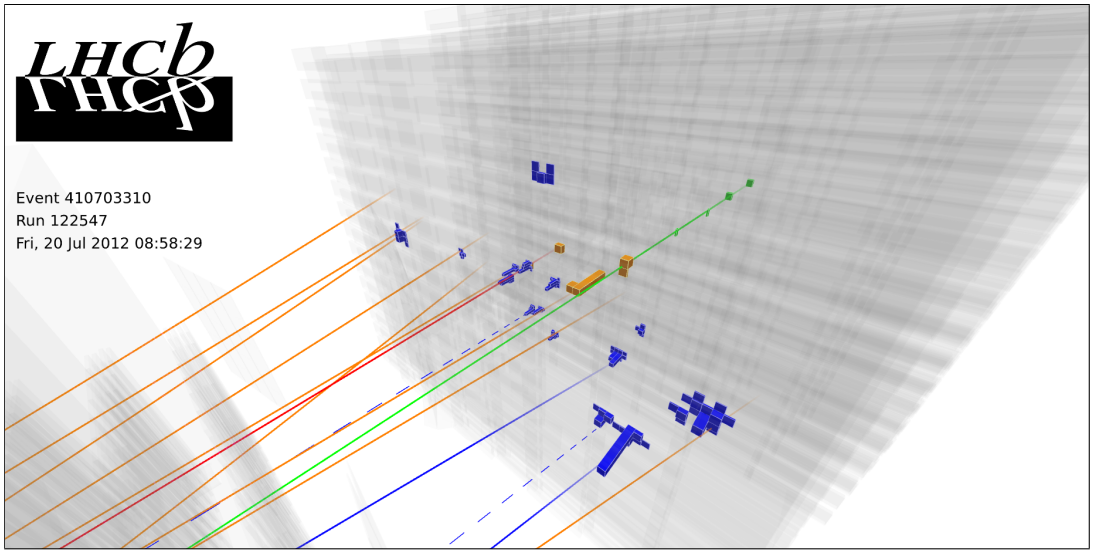


Figure 5.9: Event display, with a zoom in the VELO in the xy-plane, of the event with the most signal-like $B_s^0 \to e^{\pm}\mu^{\mp}$ candidate, number 410703310 in run 122547. The PV and secondary (decay) vertex are denoted by the large and small circle respectively.

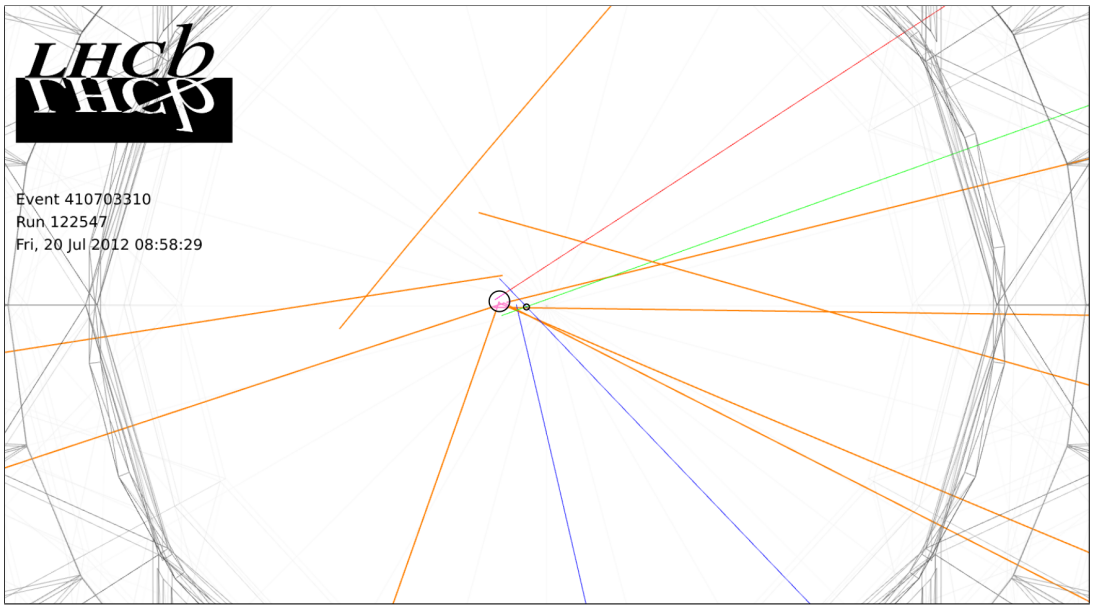


Table 5.3: Some of the most relevant variables from reconstruction of the event with the most signal-like $B_s^0 \to e^\pm\mu^\mp$ candidate, number 410703310 in run 122547, *i.e.* the event in figs. 5.6 to 5.9.

Observable	Value	Unit
$e^\pm$		
IP	0.29	mm
χ^2_{IP}	272.36	
$\chi^2_{\text{track}}/n_{\text{DOF}}$	0.72	
p	26.581	GeV/c
p_{T}	4.785	GeV/c
p_{brem}	3.718	GeV/c
`DLLe`	10.14	
`ProbNNk`	0.000	
`L0Electron`	`TOS`	
`Hlt1TrackAllL0`	`TOS`	
$\mu^\mp$		
IP	0.19	mm
χ^2_{IP}	103.94	
$\chi^2_{\text{track}}/n_{\text{DOF}}$	0.55	
p	37.199	GeV/c
p_{T}	4.355	GeV/c
`ProbNNmu` $\cdot (1 -$ `ProbNNk`$) \cdot (1 -$ `ProbNNp`$)$	0.980	
`L0Muon`	`TOS`	
`Hlt1TrackAllL0`	`TOS`	
$B^0_{(d/s)}$ $[e^\pm\mu^\mp]$		
BDT	0.906	
$\chi^2_{\text{vertex}}/n_{\text{DOF}}$	1.58	
χ^2_{IP}	0.32	
$\tau_{B^0_s}$	0.79	ps
χ^2_{FD}	676.59	
FD	2.81	mm
$m_{e^\pm\mu^\mp}$	5359	MeV/c^2
`Hlt2Topo2BodyBBDT`	`TOS`	
`Hlt2B2HH`	`TOS`	

Chapter 6

Results

To interpret the fit results shown in chapter 5, as no signal strength is measured, a limit on the branching fractions can be set. How this is done is discussed in section 6.1 and its application to the $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ branching fractions is discussed in section 6.2. The limits set by this analysis have been used by theorists to constrain new physics models. A brief summary of its theoretical application and interpretation is presented in section 6.3 , including an outlook on what can be expected in terms of precision for future measurements, in particular of LHCb.

6.1 Limit setting: the CL_s method

At the heart of this analysis lies the question: can the background-only hypothesis, *i.e.* the null hypothesis, be rejected in favour of the signal-plus-background hypothesis, depending on the values of the branching fractions of $B^0_{(d/s)} \to e^\pm\mu^\mp$? In case no significant excess is found, then another question becomes more relevant: for which values of the branching fractions, *i.e.* the to-be excluded values, can the signal-plus-background hypothesis be rejected in favour of the background-only hypothesis? To answer this question, a variable, *i.e.* a test statistic, has to be constructed that discriminates the two hypotheses. Based on the distribution of the test statistic for each hypothesis, one can determine probabilities of the observed value of the test statistic for each possible value of the branching fractions. Subsequently, confidence levels can be set.

The most powerful test-statistic is the likelihood ratio defined by,

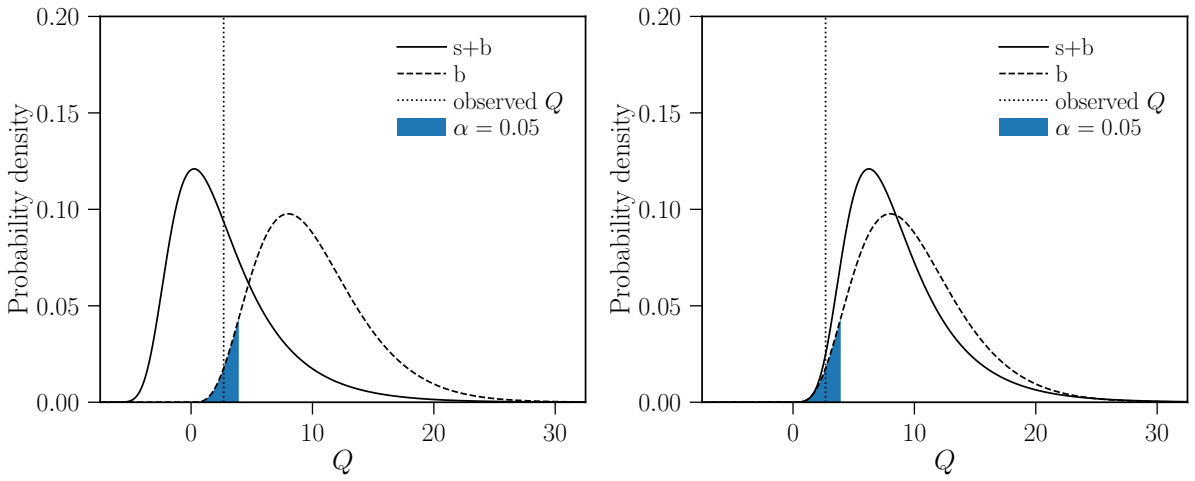
$$\Lambda = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b} \ , \tag{6.1}$$

where $\mathcal{L}_{s+b}$ and $\mathcal{L}_b$ are the likelihoods from eq. (5.1) for signal-plus-background and background only, respectively [55]. Such a likelihood ratio can intuitively be seen as how much times it is more likely that this dataset contains signal instead of only background. To use and interpret this test statistic, it is more convenient to transform the likelihood ratio to,

$$Q = -2\ln(\Lambda) \ , \tag{6.2}$$

as for large sample sizes, Q is distributed like a χ^2 distribution if the background hypothesis is true [74]. The number of degrees of freedom of this χ^2 distribution is the difference of

Figure 6.1: Example of distributions of Q for background and signal-plus-background in a case of large (left) and small (right) separation. A dashed vertical line shows a possible observed value of Q that, in case of $\alpha = 0.05$, leads to rejecting the background hypothesis in favour of signal-plus-background if one uses a left-sided p-value background rejection. In the case of bad separation this is undesirable, as the probability of a type I error is high.



degrees of freedom between the signal-plus-background hypothesis and the background hypothesis.

As noted earlier, the hypothesis test depends on the value of the branching fractions of $B^0_{(d/s)} \to e^{\pm}\mu^{\mp}$. The test has lower sensitivity at lower values of the branching fractions, as signal-plus-background is practically the same as background only. To illustrate this behaviour, two examples of separation of distributions of Q, one favourable and one unfavourable, are shown in fig. 6.1. This shows that in case of little separation, an observed left-sided p-value lower than 5% can also mean that one should reject the signal-plus-background hypothesis and therefore this test says likely more about the background than the signal! Consequently, to not make a type I error at these low values, just a left-sided p-value for the background-only distribution of Q does not suffice. To alleviate this problem, a *modified* frequentist confidence level is used by defining,

$$\mathrm{CL}_s = \frac{p_{s+b}}{1 - p_b} \ , \tag{6.3}$$

where p_{s+b} and p_b are the right-sided p-values of the signal-plus-background and the background-only hypothesis respectively [75]. These p-values are given by,

$$p_{s+b} = \int_{Q_{\mathrm{observed}}}^{\infty} f_{s+b}(Q)\, dQ \qquad\qquad 1 - p_b = \int_{-\infty}^{Q_{\mathrm{observed}}} f_b(Q)\, dQ \tag{6.4}$$

where f_{s+b} and f_b are the distribution of Q for signal-plus-background and background only respectively. With the CL_s method, the convention is to set the limit by,

$$1 - \mathrm{CL}_s = \mathrm{CL} \ , \tag{6.5}$$

where CL is the confidence level. The use of this statistic to set confidence levels is called the CL_s method.

Note that CL_s is not a pure confidence level, but a ratio of confidence levels. While deviating from a true confidence level might seem unwanted, a larger problem is the hypothesis one is actually testing if one keeps with a true confidence level. A confidence level about signal-plus-background says more about the whole dataset, while in fact the goal is to get the confidence level of only signal, since the goal is to reject the signal hypothesis at certain values of the branching fraction. As determining signal-only confidence levels is not possible if one has backgrounds, an approximation is made with CL_s. For cases with large separation, CL_s becomes a true confidence level of $s + b$, but in cases of largely overlapping distributions, the CL_s statistic is not a true confidence level, but takes into account the loss in discriminating power. This deviation does not cause a problem, since CL_s is slighlty more conservative with respect to what a true confidence level would be [75].

The best way to further clarify and illustrate the interpretation of CL_s and the limit setting with the CL_s method in general is by applying it to the actual analysis, which will be done in section 6.2.

Practically, the CL_s method is usually performed by determining the distributions of Q for a given value of the branching fraction numerically. Values of Q are generated for a discrete set of values of the branching fraction by generating pseudo-experiments from the PDF, both with signal-plus-background and only background.

6.2 Results

Before applying the CL_s method to $B^0_{(d/s)} \to e^\pm\mu^\mp$, some subtleties in the definition of the signal and background for the hypothesis testing have to be clarified. Since the expected distributions of the invariant mass of $B^0 \to e^\pm\mu^\mp$ and $B^0_s \to e^\pm\mu^\mp$ have overlap, the individual limit setting is not independent. In principle, one could see these branching fractions as two separate limits, where one defines the null hypothesis as solely Standard-Model like and the alternative hypothesis as containing one decay, either $B^0 \to e^\pm\mu^\mp$ or $B^0_s \to e^\pm\mu^\mp$. To be more general, a slightly more conservative method is chosen by using the central value of the fit result to fix the other decay.

For the limit on the $B^0_s \to e^\pm\mu^\mp$ decay, another subtlety has to be considered. As explained in section 1.2, due to mixing of the B^0_s and $\overline{B}^0_s$ flavour eigenstates and the large lifetime difference between the heavy and light mass-eigenstate, the *CP*-even and *CP*-odd eigenstates have different effective lifetimes. Since a decay could occur through a *CP*-odd, *CP*-even eigenstate, or a mixture, the CL_s method has to be performed for the two extremes, purely *CP*-odd or *CP*-even. As main result, it is chosen to use the heavy-eigenstate, *i.e.* *CP*-odd, contribution, as it is expected that the amplitude from neutrino oscillations is dominated by *CP*-odd, in a similar way as for the $B^0_s \to \mu^+\mu^-$ decay [50,76]. Nonetheless, it should be kept in mind that for any new physics contribution, it is unknown what the *CP* nature of the decay is.

The CL_s method was performed by generating pseudo-experiments with the `RooStats` software package [77]. Its results are shown in fig. 6.2. What is shown here is the observed CL_s with the black line and no-signal CL_s with the dashed black line. The green (1σ) and yellow (2σ) bands denote the distribution of the expected CL_s. For the $B^0_s \to e^\pm\mu^\mp$, it is clearly visible that the observed CL_s touches the 1σ deviation band, caused by the event shown in section 5.4. The upper limits are set where, given the branching fraction, the

Figure 6.2: Results of the CL_s method for $B^0 \to e^\pm\mu^\mp$ (left) and $B_s^0 \to e^\pm\mu^\mp$ (right) [75]. The black line represents the observed CL_s value and the dashed line the no-signal CL_s value, *i.e.* for $\mathcal{B}(B^0_{(d/s)} \to e^\pm\mu^\mp) = 0$. The green ($1\sigma$) and yellow ($2\sigma$) bands denote the no-signal distribution of CL_s. The limits are set at the observed $CL_s = 0.05$. For $B_s^0 \to e^\pm\mu^\mp$, a heavy-eigenstate contribution is assumed.

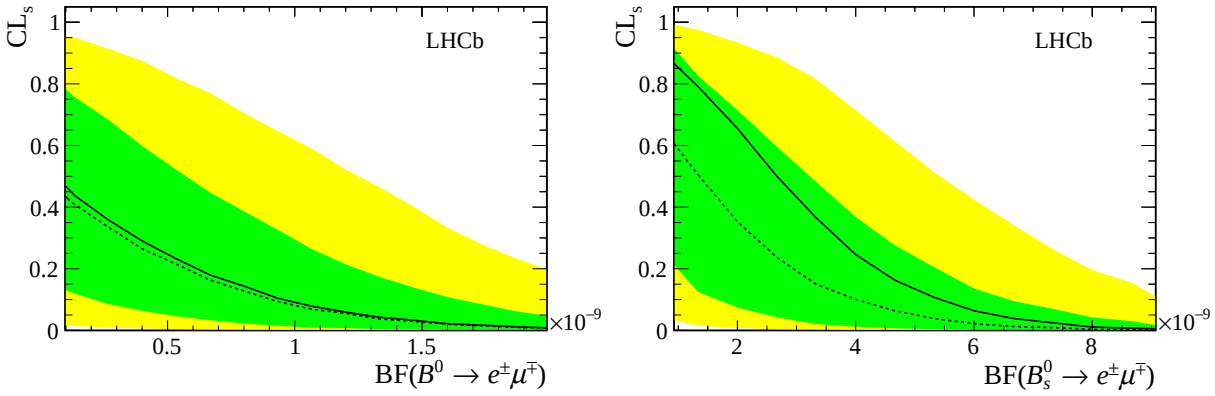


Table 6.1: Upper limits, observed and expected (assuming no signal), for $\mathcal{B}(B^0 \to e^\pm\mu^\mp)$ and $\mathcal{B}(B_s^0 \to e^\pm\mu^\mp)$ at 95% (90%) confidence level determined with the CL_s method. For $B_s^0 \to e^\pm\mu^\mp$, the main result is the *CP*-odd contribution, *i.e.* the heavy eigenstate.

Decay channel		Observed	Expected
$B_s^0 \to e^\pm\mu^\mp$	*CP*-odd	$6.3\,(5.4) \times 10^{-9}$	$5.0\,(3.9) \times 10^{-9}$
	CP-even	$7.2\,(6.0) \times 10^{-9}$	$5.6\,(4.5) \times 10^{-9}$
$B^0 \to e^\pm\mu^\mp$		$1.3\,(1.0) \times 10^{-9}$	$1.2\,(0.9) \times 10^{-9}$

confidence in the signal, *i.e.* the CL_s, goes below 5% (10%) for 95% (90%) confidence level respectively. The resulting upper limits are summarised in table 6.1, in case of $B_s^0 \to e^\pm\mu^\mp$ both for a *CP*-odd and a *CP*-even decay.

The effect of the systematic uncertainties, included in the likelihood with Gaussian priors, as shown in eq. (5.1), is about 5% on the limits.

These results improve the previous best limits (set at LHCb with 2011 data) by a factor of 2 to 3 [31]. This improvement is first of all due to a three times larger dataset and partly to the increased performance of the multivariate classifier and trigger selection.

6.3 Interpretation and outlook

Since the above described result is the world's best limit, it has been and continues to be used in many theoretical and phenomenological studies to derive constraints on new physics. Due to the connection between lepton-flavour violation and lepton-flavour universality, mentioned in chapter 1, this result has been used in model-independent analyses to put contraints on Effective Field Theory couplings related to B-anomalies [78]. In addition, it has helped to reduce the parameter space of specific models, also related to the B-anomalies, like leptoquarks [79–81]. The result has also been used by more general new physics contraints [82, 83]. Specific models not related to the B-anomalies have used

the results as well: a model for a sterile neutrino, which tempts to solve the origin of the mass of neutrinos [84]; a model for a generic new scalar [85]; and a paper studying flavour-changing neutral currents related to the Higgs [86].

Since the publication of the results obtained in this analysis dissertation, the LHCb experiment has accumulated more data on tape, namely $6\,\text{fb}^{-1}$ at a higher $\sqrt{s}$ (and thus about two times higher cross-section for b-hadrons [87]). This results in four times more b-hadrons than in this dataset. As the analysis is dominated by statistical uncertainties, the search can now be performed with a sensitivity two times higher.

In the future, LHCb will collect more data with a five times higher instantaneous luminosity and a new upgraded detector [88]. The goal is to collect $50\,\text{fb}^{-1}$. After a further upgrade, the goal of which is to go to an even higher luminosity, $300\,\text{fb}^{-1}$ will be collected [89]. Assuming similar performance in background rejection and signal efficiency, the first upgrade, starting data taking from 2021, will be able to push down the upper limits for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ to 2×10^{-10} and 8×10^{-10}, respectively. Under the same assumption for the second upgrade, the upper limits for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ will go down to 9×10^{-11} and 3×10^{-10}, respectively. LHCb will therefore remain the leading experiment in the search for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$.

Part III

Performance of electron reconstruction at LHCb

Introduction to Part III

The use of electrons in analyses at LHCb has taken a more prominent role because of the anomalies seen in the measurements of lepton-universality ratios, discussed in chapter 1.

Compared to other particles, electrons are more difficult to detect at LHCb. Due to their low mass, both bremsstrahlung from material and final-state radiation is larger with respect to muons, pions, kaons and protons. This makes the electron reconstruction more sensitive to the detector description in simulation and less efficient, and the electron momentum more difficult to infer. As the statistical uncertainties will go down for future analyses of LHCb, a better understanding is needed of how electrons behave in the detector to reduce systematic uncertainties. For this purpose, a new method has been developed to measure their reconstruction efficiencies in data.

To give context to this method, the behaviour of electrons and how they are treated at LHCb will be discussed. First, the physics behind bremsstrahlung (their main energy-loss mechanism) and final-state radiation will be discussed in chapter 7. After travelling through the detector, electrons are reconstructed with general reconstruction algorithms for charged particles. These and other algorithms to reconstruct additional objects, like bremsstrahlung photons, are discussed in chapter 8. In chapter 9, the study that lead to the novel method to determine electron-reconstruction efficiencies in data will be discussed.

Chapter 7

Physics of bremsstrahlung

Electrons are more affected by energy loss at energy scales relevant to LHCb than other charged particles. The main energy-loss mechanism is bremsstrahlung, which is in essence electromagnetic radiation coming from an accelerating charged particle. For charged particles originating from decays, photon emmision can happen at two levels: emitted in a decay that produced the charged particle, discussed in section 7.1, or emitted by interactions with charged particles in a medium, *i.e.* while travelling through the detector, discussed in section 7.2.

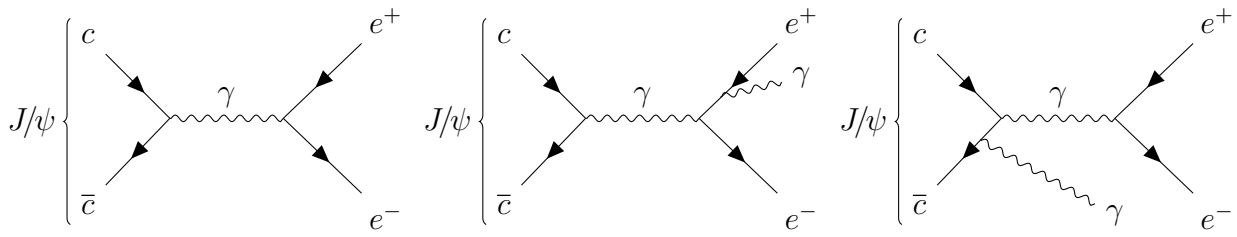
7.1 Radiative losses in decays

Photon emission in decays can be divided in two categories: initial-state and final-state radiation, depicted for the decay of $J/\psi \rightarrow e^+e^-$ in fig. 7.1. While initial-state radiation depends on the decay, final-state radiation can be treated much more generically.

To illustrate the decay-dependent behaviour of initial-state radiation, the decay of $J/\psi \rightarrow e^+e^-$ is compared to $\pi^0 \rightarrow e^+e^-$. Since the leptonic decay channels of the J/ψ are dominated by the electromagnetic interaction, which conserves C-parity, and the J/ψ has a charge-parity eigenvalue of -1 (same as the photon), the decay $J/\psi \rightarrow \gamma\gamma$ is forbidden. Hence, also the decay of $J/\psi \rightarrow e^+e^-\gamma$, where the e^+e^- originates from a virtual photon, is forbidden. While the decay of the π^0 is dominated by the electromagnetic interaction as well, its charge-parity eigenvalue is $+1$, hence it is dominated by $\pi^0 \rightarrow \gamma\gamma$. Accordingly, the decay of $\pi^0 \rightarrow e^+e^-\gamma$ occurs about 2×10^5 times more than $\pi^0 \rightarrow e^+e^-$ [16]. This clearly shows that a decay with or without photons can behave very different, depending on the type of decay. It should be noted that due to available phase space for initial-state radiation, it is more likely that initial-state photons are not soft. This is crucial in comparing initial-state to final-state radiation.

Contrary to initial-state radiation, the behaviour of final-state radiation can be understood in a general way. Consider a decay with outgoing, real, charged particles. These particles can have photon loops in higher-order QED corrections to the decay width, which lead to an infinite amplitude. These infinite virtual corrections are cancelled by those of real photons, *i.e.* external photon lines. These real photons can be considered part of the decay, since adding very soft photons can be indistinguishable with respect to the decay without these photons. Hence, these non-virtual photons add to the total amplitude and are collectively called final-state radiation and sometimes bremsstrahlung as well [8].

Figure 7.1: Leading-order Feynman diagrams of (from left to right) $J/\psi \to e^+e^-$, $J/\psi \to e^+e^-\gamma$ (final-state radiation) and $J/\psi \to e^+e^-\gamma$ (initial-state radiation). The latter is forbidden as the electromagnetic interaction is charge-parity conserving.



The quantitative behaviour of final-state radiation can be understood from adding an external photon line to a charged final-state particle in the first-order Feynman diagram of a decay. This leads to the following correction to the first-order differential cross-section $\mathrm{d}\sigma_0$,

$$\mathrm{d}\sigma_{\mathrm{FSR}} = \mathrm{d}\sigma_0 \; f(k, \cos\theta, \phi) \; k \; \mathrm{d}k \; \mathrm{d}\cos\theta \; \mathrm{d}\phi \; , \tag{7.1}$$

where $\mathrm{d}\sigma_{\mathrm{FSR}}$ is the corrected cross-section, k is the photon energy, θ and ϕ are the polar and azimuthal angle between the charged particle radiating the photon and the photon itself in the rest-frame of the decaying particle without the photon, respectively [90]. The factor $f(k, \cos\theta, \phi)$ is given by,

$$\begin{aligned} f(k, \cos\theta, \phi) = & \frac{(2\pi)^2}{m_P^2} \left[\delta\left(\frac{k}{m_{\mathrm{P}}}\right) \frac{m_{\mathrm{P}}}{k} \left(1 - \frac{\alpha}{\pi} N\right) + \Theta(k - \epsilon) \frac{\alpha}{\pi} \left(1 + \left(1 - \frac{k}{2k_{\max}}\right)^2\right) \right. \\ & \left. \times \frac{m_{\mathrm{P}}^2}{2k^2} \frac{2}{1 - \cos\theta \sqrt{1 - m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2 / (m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2}} \right] , \end{aligned} \tag{7.2}$$

where m_{P} is the mass of the decaying particle and m_{ch} is the mass of the charged final-state particle radiating the photon; the maximum photon energy is $k_{\max} = m_{\mathrm{P}}/2 - (m_{\mathrm{ch}} + m_{\mathrm{Y}})^2/(2m_{\mathrm{P}})$; ϵ is the cut-off energy to get rid of IR divergence; and α is the fine-structure constant. The factor N for the soft-photon term is given by,

$$\begin{aligned} N = & \frac{1}{\sqrt{1 - m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2 / (m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2}} \ln \left(\left(1 + \sqrt{1 - \frac{m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2}{(m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2}} \right)^2 \frac{(m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2}{m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2} \right) \\ & \times \left(\ln \frac{k_{\max}}{\epsilon} - \frac{3}{4} + \frac{\epsilon}{k_{\max}} - \frac{\epsilon^2}{4k_{\max}} \right) . \end{aligned} \tag{7.3}$$

The take-away message from eqs. (7.1) to (7.3) is their dependence on m_{P}, making it decay-specific, and m_{ch}, which makes it dependent on the mass of the final-state particle.

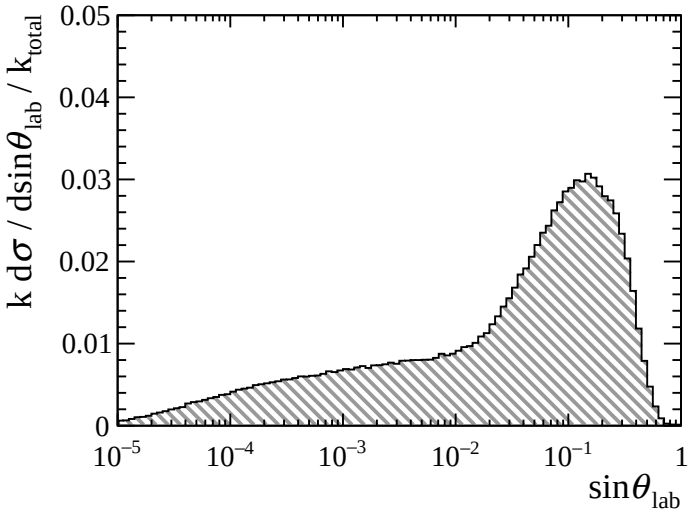
The radiated energy distribution, differential in k and $\cos\theta$, for energies above the IR cut-off, is given by,

$$k \; \mathrm{d}\sigma_{\mathrm{FSR}} = \mathrm{d}\sigma_0 \; 8\pi^2 \alpha \left(1 + \left(1 - \frac{k}{2k_{\max}}\right)^2\right) \frac{2\mathrm{d}k \; \mathrm{d}\cos\theta}{1 - \cos\theta \sqrt{1 - m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2 / (m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2}} \; . \tag{7.4}$$

By integrating over k and $\cos\theta$, it can be shown that the total amount of radiated energy is proportional to,

$$\int k \mathrm{d}\sigma_{\mathrm{FSR}} \propto \sigma_B k_{\max} \frac{\tanh^{-1}(\sqrt{1 - m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2 / (m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2})}{\sqrt{1 - m_{\mathrm{ch}}^2 m_{\mathrm{P}}^2 / (m_{\mathrm{ch}}^2 + m_{\mathrm{P}}^2)^2}} \; , \tag{7.5}$$

Figure 7.2: Distribution of radiated energy (in the e^+e^- rest frame, normalised with respect to the total radiation energy k_{total}) from final-state radiation of the $J/\psi \to e^+e^-$ decay, $k\mathrm{d}\sigma/k_{\text{total}}$, as a function of the sine of the angle between the lost momentum and one of the electrons, $\sin\theta_{\text{lab}}$ (in the lab frame), in simulation of the decay $B^+ \to J/\psi(\to e^+e^-)K^+$. The final-state radiation is simulated with PHOTOS [46]. Note the log-scale for the $\sin\theta_{\text{lab}}$-axis.



which shows that the amount of final-state radiation approximately rises with $\log(m_{\text{P}}/m_{\text{ch}})$ if $m_{\text{P}} \gg m_{\text{ch}}$. Hence, the smaller the final-state particle mass, the more final-state radiation. Using eq. (7.5) for a decay of $J/\psi \to e^+e^-$ compared to $J/\psi \to \mu^+\mu^-$, the amount of radiated energy is about 2.3 times higher.

The median of the angular distribution of θ, θ_0^{FSR}, is given by,

$$\cos\theta_0^{\text{FSR}} = \frac{1 - m_{\text{ch}}m_{\text{P}}/(m_{\text{ch}}^2 + m_{\text{P}}^2)}{\sqrt{1 - m_{\text{ch}}^2 m_{\text{P}}^2/(m_{\text{ch}}^2 + m_{\text{P}}^2)^2}} \,, \tag{7.6}$$

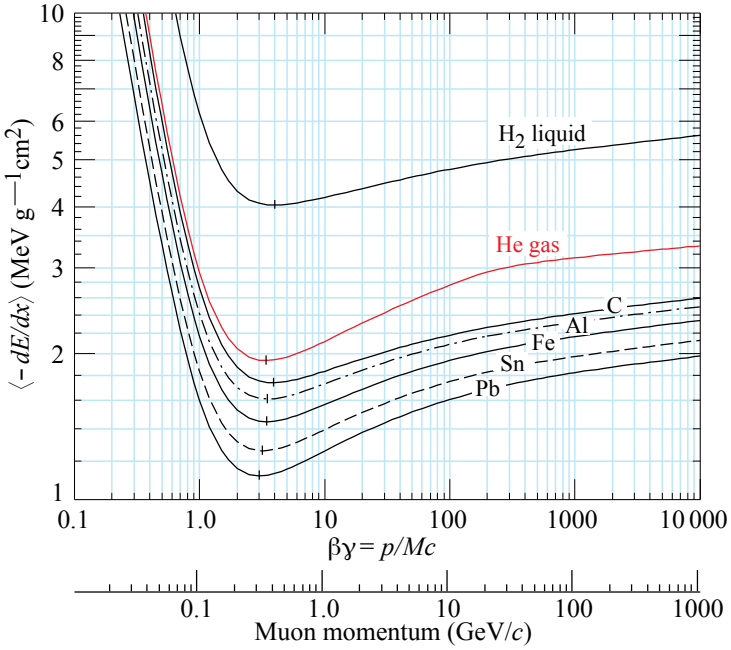
calculated with eq. (7.4). For an electron in the decay of $J/\psi \to e^+e^-$, this results in a θ_0^{FSR} of 0.9°, which in general is even smaller in the lab frame. Therefore, final-state radiation is emitted essentially parallel to the charged final-state particles. This notion will be important for comparisons with material interactions.

To account for final-state radiation in simulation of decays at LHCb, the algorithm PHOTOS is used, which uses eq. (7.1) [46, 47]. To illustrate the effect final-state radiation has on a typical decay involving electrons at LHCb, the angular distribution of the radiated energy (in the lab frame) for electrons from the $B^+ \to J/\psi(\to e^+e^-)K^+$ decay is shown in fig. 7.2.

7.2 Material interactions

After decay, charged final-state particles travel through the detector and interact electromagnetically with the electrons and nuclei of the detector material. In case of hadrons, final-state particles interact through the strong interaction with the nuclei as well, also for neutral final-state particles. While this is the basis of detecting these particles in the first place, it also has important degrading effects on the inference of its momentum, as these interactions can cause energy loss and deflections.

Figure 7.3: Mass stopping power for a range of different materials in the range of the Bethe-Bloch equation, eq. (7.7). Picture adapted from [16].



For heavy charged relativistic particles, the energy loss while travelling through a material is dominated by ionisation. The mean of this energy loss per unit of density per unit of distance (called mass stopping power) for a particle with unit charge in the range of $\beta\gamma$, *i.e.* speed in units of c times the Lorentz factor, from 0.1 to 1000 is well described by the Bethe-Bloch formula,

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = K \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 \beta^2 \gamma^2 W_{\max}}{I} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] , \tag{7.7}$$

where Z and A are the atomic number and mass of the material, respectively; $K = 4\pi N_\mathrm{A} r_e^2 m_e^2 c^2$ and therefore is a constant, *i.e.* not dependent on the material; $W_{\max}$ is the maximum amount of energy that can be transferred; I is the mean excitation energy of the material; $\delta(\beta\gamma)/2$ is the density-effect correction, which is proportional to $\ln(\beta\gamma)$ at high $\beta\gamma$; and m_e and r_e are the electron mass and classical radius, respectively [16]. This shows that the mass stopping power rapidly drops up to about $\beta\gamma = 1$, while for $\beta\gamma > 1$ the energy loss increases slowly with $\ln(\beta\gamma)$, as shown in fig. 7.3. Thus, a minimum in the energy loss occurs at $\beta\gamma$ of $\mathcal{O}(1)$. A particle at this minimum is called a minimum ionizing particle (MIP). For many solids and gases the mass stopping power for a MIP is in the range of 1 to 2 MeV cm^2/g. Compared to the mean of the energy loss distribution, the variance is small enough to be irrelevant for the precision at LHCb [91].

For charged particles that are reconstructed in the LHCb detector, the relevant energy range starts at a few GeV and that means that for muons, pions, kaons and protons $\beta\gamma$ stays below 10^3 up to energies of 100 GeV. Accordingly, these particles are considered MIPs, as the energy loss rises only logarithmically with energy. For electrons, because of their small mass, $\beta\gamma$ starts at $\mathcal{O}(10^3)$ and therefore falls outside the range of the Bethe-Bloch equation, where material bremsstrahlung starts to dominate.

Material bremsstrahlung is essentially a charged particle scattering from a virtual

photon of the Coulomb field of a nucleus of the material with the addition of an outgoing real photon. In a more classical description, it is radiation by a charge decelerating in the Coulomb field of the nucleus. For highly relativistic charged particles, it can be shown that the cross-section, σ, of producing bremsstrahlung photons is of the order,

$$\sigma \sim \alpha Z^2 \left(\frac{e^2}{mc^2} \right)^2 , \tag{7.8}$$

where m is the mass of the particle and Z is the nuclear charge [92]. This shows that the cross-section is proportional to m^{-2} and therefore the electron, having such a small mass, is much more affected than other particle species.

The mean energy loss by bremsstrahlung for electrons (in terms of the linear stopping power in units of energy per distance) is given (for energies higher than 1 GeV) by,

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{E}{X_0} , \tag{7.9}$$

where X_0, the radiation length, is given by,

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A}{A} \left[Z^2 (L_{\mathrm{rad}} - f(Z)) + Z L'_{\mathrm{rad}} \right] , \tag{7.10}$$

where L_{rad} and L'_{rad} are radiative corrections depending logarithmically on Z and $f(Z)$ is a small correction due to screening by atomic electrons of the nuclear Coulomb field [16, 93]. Combining bremsstrahlung with ionisation and other smaller contributions, the energy loss of electrons, depending on its energy, is depicted in fig. 7.4. It shows that at energies relevant for LHCb, bremsstrahlung is the only relevant energy loss.

While for ionisation the energy loss variance is rather small compared to the mean, this is not the case for bremsstrahlung. The intensity distribution (proportional to probability times energy) is approximately constant with respect to bremsstrahlung energy as it is given by,

$$k \mathrm{d}\sigma = \frac{A}{X_0 N_{\mathrm{A}}} \left[\frac{4}{3} - \frac{4}{3} y + y^2 \right] \mathrm{d}k \tag{7.11}$$

assuming energies below 1 TeV and where $y = k/E$ [16, 94, 95]. The resulting distribution of how much of the fraction of energy, $z = 1 - y$, of the initial particle is left can be obtained by a rough but convenient approximation of eq. (7.11) by,

$$k \mathrm{d}\sigma \propto \frac{y}{\ln(1/(1-y))} \mathrm{d}k , \tag{7.12}$$

resulting in the distribution of z,

$$f(z) = \frac{[-\ln(z)]^{c-1}}{\Gamma(c)} , \tag{7.13}$$

where Γ is the Gamma function and $c = x/(X_0 \ln(2))$ [92]. The distributions of the intensity and of z are shown in fig. 7.5. This shows that the spread of energies of electrons after traversing different amounts of material is large and differs significantly per thickness.

Figure 7.4: Energy loss per total energy per unit of radiation length of electrons in lead, as a function of the electron's energy. For electrons at LHCb, relevant energies start at 1 GeV, so energy losses are dominated by bremsstrahlung. Picture from [16].

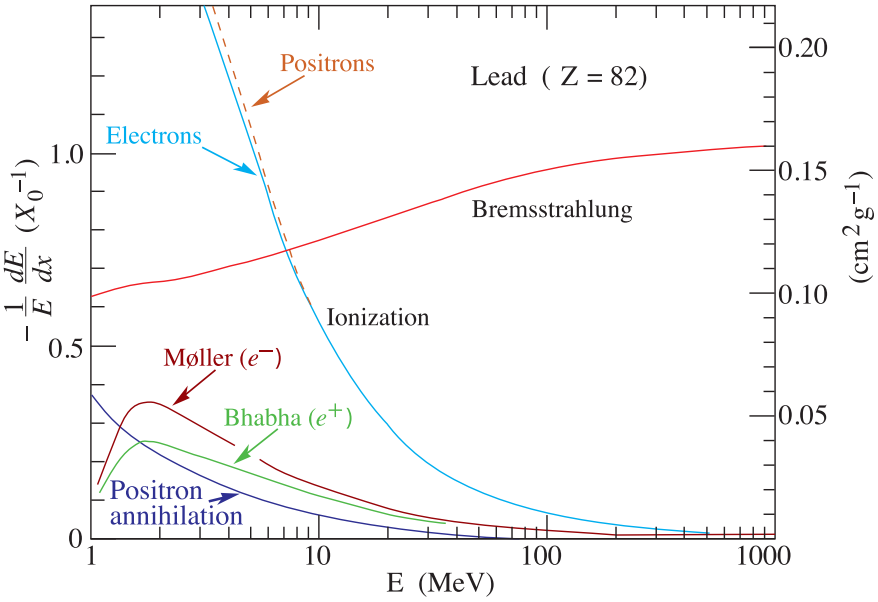
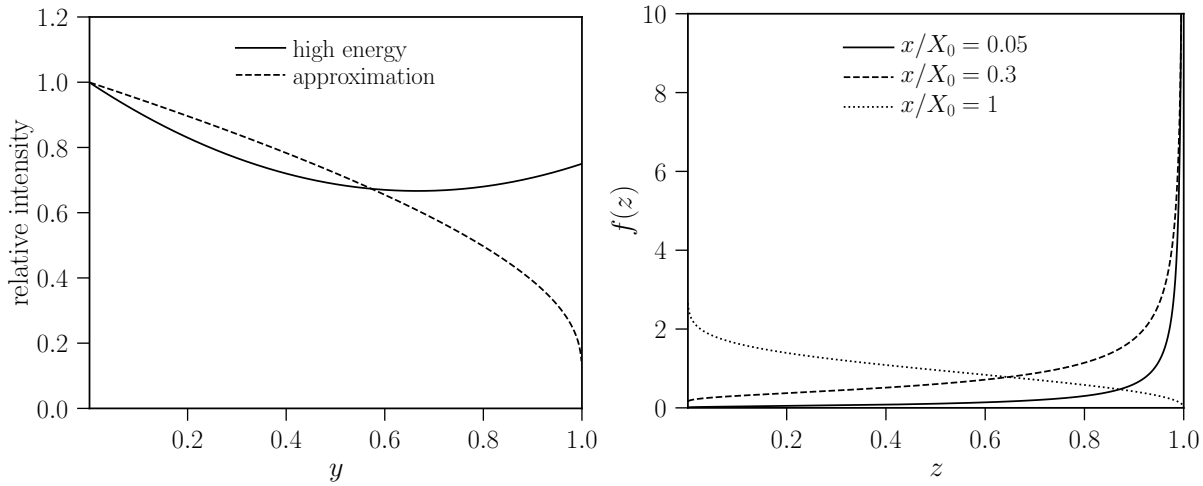


Figure 7.5: To the left: distribution of relative intensity, normalised with respect to the value at $y = 0$, from an electron by bremsstrahlung. The solid line and the dashed line denote eq. (7.11) and the approximation used to derive eq. (7.13), respectively. To the right: distribution of the fraction of remaining energy of an electron, $f(z)$, from bremsstrahlung with different levels of thickness of traversed material in units of radiation length.



It should be noted that beside bremsstrahlung caused by material interactions, it can also originate from the charged particle bending in a magnetic field. The energy loss can be estimated using the relativistic generalisation of the Larmor formula [96]. Assuming it is highly relativistic, an electron travelling through the LHCb magnet, which has a bending power of 4 Tm, will loose less than 0.1% of its energy below 200 GeV. Subsequently, bremsstrahlung from the LHCb magnetic field is negligible.

As an example of energy loss in general for LHCb, consider a pion and an electron, with a typical energy of 20 GeV for b and c hadron decays, travelling through 5 cm of aluminium, which is of the order of the material encountered in the LHCb tracking stations. In this case, the mean energy loss for the pion by ionisation, assuming a stopping power of 1.5 MeV cm^2/g, is about 20 MeV, *i.e.* 0.1% of its energy. For the electron, the mean energy loss is about 8 GeV, *i.e.* 39% of its energy, given a radiation length of 8.9 cm [93].

Apart from energy losses, another type of material interaction is important that does not cause large energy losses, but does cause significant angular deflections. For relativistic heavy charged particles, Coulomb scattering from nuclei is the dominant cause of deflections. The scattering occurs mainly through many small-angle scatters, therefore called multiple scattering, which, in accordance with the Central Limit Theorem, behaves Gaussian. The RMS of the distribution of the polar scattering-angle for a particle of unit charge is given by,

$$\theta_0 = \frac{13.6\,\text{MeV}}{\beta cp}\sqrt{\frac{x}{X_0}}\left[1 + 0.038\ln\left(\frac{x}{X_0\beta^2}\right)\right] , \tag{7.14}$$

where p is the momentum of the particle and x is the distance travelled through the material [16]. For energies relevant for LHCb, this effectively means that $\theta_0 \propto p^{-1}$. Note that this holds for both electrons and MIPs given energy. Despite that, multiple scattering of electrons is effectively higher as the electron looses more energy than other particle species integrated over the trajectory. For a charged relativistic particle with an energy of 20 GeV, this means a θ_0 of 0.5 mrad after traversing 5 cm of aluminium.

Next to the deflection of a charged particle itself, the angular distribution of bremsstrahlung is also relevant. The characteristic angle of emission with respect to an outgoing electron, θ_0^{brem}, for bremsstrahlung photon emission is,

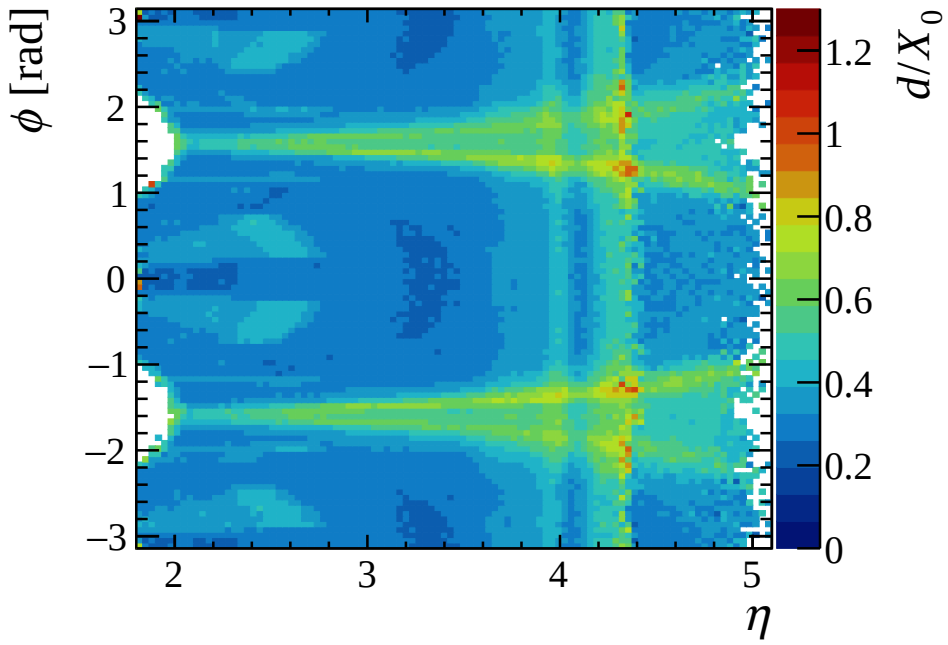
$$\theta_0^{\text{brem}} = \frac{m_e c^2}{E} , \tag{7.15}$$

where E is energy of the electron and m_e is the electron mass [92]. For angles larger than this, it is the electron that deflects with respect to the incident electron, the photon goes practically the same direction as the incident electron [97]. For an electron of 20 GeV, $\theta_0^{\text{brem}} = 0.026$ mrad, corresponding to a deflection of the order of 0.26 mm in the electromagnetic calorimeter, far below the size of the electromagnetic calorimeter cells. Note that, in terms of angular distributions, bremsstrahlung photons from material before the magnet are similar to final-state radiation photons, making these two cases practically indistinguishable for the experiment. Because of their similarities in general, final-state radiation is often called bremsstrahlung as well.

Material budget of LHCb

As discussed above, the amount of material in a detector has a large effect on energy loss and consequently the momentum determination of electrons. Because the momentum of charged particles is measured by their deflection in the magnetic field, the most important material budget is the material before and in the magnet. Since the volume in the magnet is not instrumented, most material is found before the magnet. The relevant material budget at LHCb in terms of radiation lengths is shown in fig. 7.6. It is integrated over the tracks of electrons from the simulation of the decay $B^0 \to J/\psi\,(\to e^+e^-)K^{*0}$, from the

Figure 7.6: Average amount of radiation lengths encountered by electrons from a $B^0 \to J/\psi(\to e^+e^-)K^{*0}$ decay in the LHCb detector from origin to the end of the TT-stations (right before the magnet) in simulation as a function of η and ϕ of the electron.



origin to the end of the TT-stations, depending on their direction in η and ϕ. The average radiation length for this sample is 0.35, which results in a mean relative energy loss of 30%. With an average momentum of about 30 GeV/c for these electrons, the mean energy loss is 9 GeV.

Chapter 8

Reconstruction

To understand the behaviour of electrons at LHCb, it is not only necessary to know how they interact with the detector, but also to understand the way they are reconstructed from the hits they leave in the detector.

Charged-particle tracks, including those from electrons, are all reconstructed with the same method. This general method will be described in section 8.1. Additional reconstructed objects can be associated to the track. The ones that are of specific importance to electrons, clusters of energy deposits in the electromagnetic calorimeter, from the track itself or from an associated bremsstrahlung photon, are described in section 8.2 and section 8.3 respectively. After this, a summary of types of electron reconstruction will be given in section 8.4. An illustration of how electrons and other reconstructed objects behave within the detector, an event display will be shown in section 8.5.

8.1 Track reconstruction

Reconstruction of charged-particle tracks consists of two steps. First, one has to find the track by associating hits in the tracking stations. Secondly, one has to determine the properties of the track, *e.g.* its momentum at the start of the track. The type of tracks and its corresponding track finding algorithms, called pattern recognition algorithms, and the track parameter determination, called the track fit, are described in section 8.1.1 and section 8.1.2 respectively.

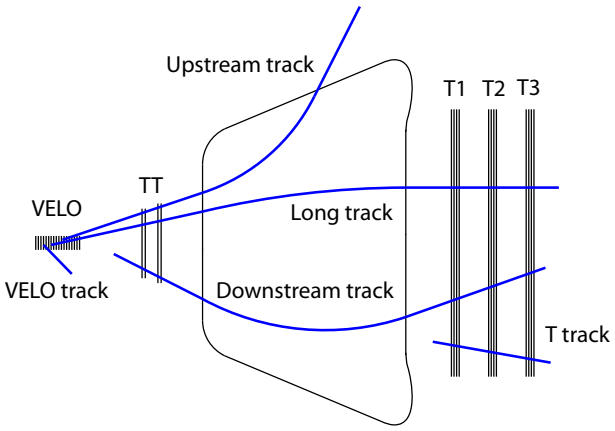
Important to note is that both with pattern recognition and the track fit, one has to keep certain performance criteria in mind: the speed of the algorithm, the efficiency of finding a track, the rejection of fake tracks and the resolution and bias in track parameters. This is especially important for the use of these or similar algorithms in the high-speed, demanding online environment of the trigger.

Additional information on track reconstruction and in particular the track fit for MIPs can be found in [91].

8.1.1 Pattern recognition

Track types are dictated by the available tracking sub-detectors. From the collision point within in the VELO, reconstructible forward tracks traverse the detector from VELO, through TT-stations, the magnet and eventually the T-stations. As the magnetic-field

Figure 8.1: Illustration of types of tracks in LHCb. Tracking sub-detectors are depicted with in the middle the magnet.



strength within the VELO and the T-stations is low, tracks within these detectors can be approximated by straight lines at first. This notion lies at the basis of the pattern recognition.

The types of tracks are illustrated schematically in fig. 8.1 and their description is listed below.

Track types

- **Long:** The main track type used in analyses. This track type is for tracks traversing the entire tracking system. It consists of at least hits in the VELO and the T-stations.
- **Downstream:** Used for long-lived particles that can decay outside the VELO, like $K^0_S \to \pi^+\pi^-$, $\Lambda^0 \to p\pi^-$ and also converted photons $\gamma \to e^+e^-$. It consists of hits in the TT and T-stations.
- **Upstream:** Used to reconstruct low-momentum tracks that bend out of the acceptance of the T-stations due to the magnet. Consists of hits in the VELO and TT-stations. At low momentum a momentum measurement is still feasible due to the fringe of the magnetic field in the TT-stations. They are of special interest for electrons, which will be discussed in section 8.4.1. Currently mainly used for RICH1 reconstruction.
- **VELO:** These tracks are mostly used to reconstruct primary vertices. Many have low pseudo-rapidity, giving them increased precision to measure a primary vertex.
- **T:** Only visible in T station. Rarely used in analyses. Mainly used for calorimeter reconstruction, essential for photon-electron separation, and RICH2 reconstruction.

The pattern recognition for each type of track is listed below chronologically, to emphasise that some of the pattern-recognition output is used in subsequent pattern-recognition algorithms.

Pattern recognition

- **VELO seeding:** The algorithm, called `FastVelo`, is based on the notion that tracks are approximately straight lines in the VELO [98]. A VELO track is defined by having at least three r and three ϕ sensor hits. For forward tracks, starting from the last four sensors, that have the most separation, four r-sensor hits are searched for that form a straight line in the rz projection. With the remaining hits the procedure is repeated, requiring three r-sensor hits. Extra hits are added going upstream. Subsequently, ϕ sensor hits are added to form full 3D straight lines. To improve efficiency, especially for non-pointing tracks originating from long-lived particles, an additional track finding algorithm is applied on the remaining hits. The resulting VELO tracks serve as seeds for subsequent pattern recognition algorithms.

- **Forward:** The first algorithm searching for long tracks. It starts with VELO tracks and tries to extend these towards the T-stations. Neglecting multiple scattering, the full trajectory is known with one hit in the T-stations. Using polynomials as parametrisation of the track, the distance from the estimated trajectory towards other hits in both T- and TT-stations is calculated. The other hits are added to the track if this distance is small enough.

- **T seeding:** Remaining tracks in the T-stations are searched for by first assuming that they are straight lines. Given the low magnetic-field strength there, additional bended tracks are searched for by using a second-order polynomial. A requirement on a likelihood that takes into account expected and observed hits is set to reject fake tracks.

- **Track matching:** To improve the long tracking efficiency, T-tracks are attempted to match to the remaining VELO tracks. A requirement on matching χ^2 is set to reject fake matches.

- **Upstream:** Remaining VELO tracks are attempted to match to hits in the TT-stations, assuming strait lines in the yz plane. For low-momentum tracks, the bending in the low magnetic-field strength is still sufficient to get a momentum measurement with a resolution of 20%. Fake tracks are rejected by requiring a maximum track χ^2.

- **Downstream:** Downstream tracks are searched for by starting with remaining T-tracks and adding TT-station hits in similar fashion to forward tracking. To improve efficiency, a second method is applied by assuming that the track comes from the interaction point. This allows to estimate the momentum and in turn extrapolate towards the TT-stations, adding matching hits.

8.1.2 Track fit

After building tracks with the pattern recognition algorithms, it is the purpose of the track fit to extract the optimal parameters of the track. As these parameters change while traversing the detector due to material interactions and the magnetic field, a track

is represented by a collection of track states. A track state is given by the vector,

$$\vec{x}(z) = \begin{pmatrix} x \\ y \\ t_x \\ t_y \\ q/p \end{pmatrix}, \qquad \text{with} \qquad t_x = \frac{\partial x}{\partial z}, \qquad t_y = \frac{\partial y}{\partial z}, \tag{8.1}$$

where x, y and z are the spatial coordinates of the state, q is the charge of the track, assumed to be $\pm e$, and p is the momentum of the track state. A covariance matrix, C, is associated to each track state. The combination of a track state, $\vec{x}_k$, with a measurement, m_k, is called a node and are linked by,

$$m_k = h_k(\vec{x}_k) + \epsilon_k \ , \tag{8.2}$$

where h_k is the projection function used to link the track state to the measured quantity and ϵ_k is the measurement noise. As the track travels through the detector, the direction and its momentum changes due to the magnetic field or interactions with the material. The propagation of a track state, $\vec{x}_{k-1}$, from node $k-1$ to the track state at subsequent node k, $\vec{x}_k$, is described by,

$$\vec{x}_k = f_k(\vec{x}_{k-1}) + \vec{\omega}_k \ , \tag{8.3}$$

where f_k is the track model used to describe the track travelling through the magnetic field and $\vec{\omega}_k$ is the process noise generated by material interactions.

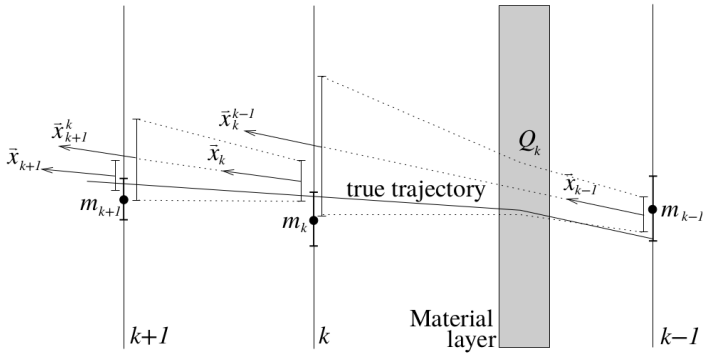
With these relations the goal of the track fit is to use the knowledge of the detector to construct a track model, estimate the process and measurement noise and combine them with the measurements, *i.e.* the hits associated to the track, to estimate the track state parameters and its uncertainties. For this purpose a Kalman Filter is used at LHCb.

Kalman Filter

The idea behind the Kalman Filter is to add measurements sequentially in such a way that it is equivalent to minimizing the χ^2 without the need to invert large matrices for the full set of measurements [99]. Apart from this computational advantage, it allows for a natural way of including process noise. The procedure is divided into the following steps: prediction, filter and smoother.

At the prediction step the goal is to estimate the track state at the next node. A first estimate of the track parameters and its covariance matrix is usually based on input from the pattern recognition. From the first estimate, the track model, f_k, is used for the propagation, including the covariance matrix estimation, which has a process noise term, Q_k, added. Due to the inhomogeneity of the magnetic field at LHCb, the track model used to describe the propagation is an adaptive fifth-order Runge-Kutta method, to solve the differential equation describing the propagation [100]. The process noise term is determined assuming the track is from a MIP, *i.e.* ionisation and multiple scattering are assumed to be the only material interactions, which is discussed in section 7.2. As the stochastic nature of multiple scattering is Gaussian and the variance of ionisation energy-losses negligible, a χ^2 minimisation and therefore a Kalman Filter is appropriate for a MIP but not necessarily for an electron. Since ionisation and multiple scattering depend

Figure 8.2: Schematic overview of sequential application of measurements with the Kalman Filter. It shows how material layers add to the estimated uncertainty in the propagation, *i.e.* the prediction step. The filtering step is shown for example by going from $\bar{x}_k^{k-1}$ to $\bar{x}_k$ due to adding m_k. Illustration from [91].



on the amount and type of material encountered, the process noise term is calculated using a material map, integrating from one node to the other.

With this track state estimate at the next node, a residual can be calculated with respect to the measurement. Using the covariance matrix, the predicted contribution to the total χ^2 can be calculated by,

$$\chi_k^2 = (m_k - h_k(\vec{x}_k))(V_k + H_k C_k H_k^T)^{-1}(m_k - h_k(\vec{x}_k)) \ , \tag{8.4}$$

where V_k, C_k and H_k are the covariance matrix of the measurement, the predicted covariance matrix at the next node and the projection matrix respectively. With the linearity of the problem, either assuming straight tracks or using a Taylor expansion, the updated total χ^2 can be minimised to update the track state parameters. The updating of a track state by adding the measurement at the node is called filtering. This process of prediction, including material effects, and filtering is illustrated in fig. 8.2.

After applying the prediction and filtering steps sequentially for all nodes, the previous track states can be updated with smoothing equations [101]. With the smoothing, an updated χ^2 can be calculated. If the contribution to the χ^2 of each node is too big, a measurement can be removed. This increases the purity, *i.e.* the track contains less hits that do not originate from the real track. If hits are removed, the track fit is repeated.

With the track fit, with its parameters and the uncertainty on the parameters, other objects and variables can be constructed like decay vertices, impact parameters, flight distances, life times and many others. Hence, the estimated uncertainties are very useful for background rejection. Especially for electrons, these estimated uncertainties are different from MIPs. This is not necessarily wrong, it is just that derived variables usually perform less well, *e.g.* in terms of background rejection. Often it is more important that they are well simulated.

Gaussian Sum Filter

With the Kalman Filter, since it is equivalent to the minimisation of the χ^2, the underlying assumption is that the measurements, energy loss and multiple-scattering angles are normally distributed. While this is approximately true for MIPs, for electrons it is not.

To alleviate this issue of non-Gaussian uncertainties in the energy loss of electrons and subsequent track fit biases, an extension of the Kalman Filter can be applied by using a Gaussian Sum Filter (GSF) [102].

The idea of a GSF is to run multiple Kalman Filters in parallel with the purpose of approximating the energy-loss distribution with a sum of Gaussians. At each prediction step, multiple Kalman prediction steps are calculated that carry each a weight according to the energy-loss distribution given by the amount of traversed material. This leads to a combinatoric explosion of the number of Kalman Filters running in parallel with increasing number of added nodes. This can be reduced by merging the ones that are most similar, subsequently keeping a maximum number of Kalman Filters running.

The use of a GSF at ATLAS and CMS have lead to improvements in the momentum resolution of electrons [103, 104]. To implement the GSF at LHCb has not been tried yet. It is expected that for LHCb the improvements are less, since most bremsstrahlung losses occur in regions where there is not a lot of magnetic field strength and different amount of bending is the information showing energy loss along the track.

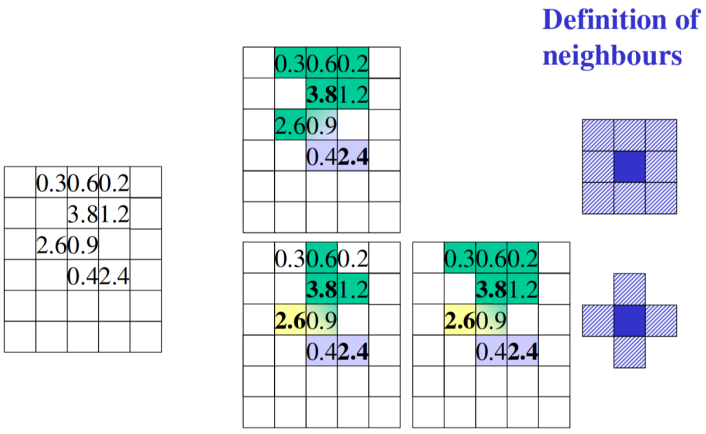
8.2 Reconstruction of showers in the calorimeter

To determine the energy and momentum of photons, *e.g.* for bremsstrahlung recovery, and aid in the identification of electrons, hadrons and muons, energy deposits in the calorimeter are reconstructed. These energy deposits, *i.e.* calorimeter showers, caused by photons, electrons and hadrons are reconstructed by clustering neighbouring electromagnetic calorimeter cells with a non-zero energy deposit, henceforth called clusters, as described in [105].

While the calorimeter consists of the SPD, PS, electromagnetic calorimeter and hadronic calorimeter, the clustering is only performed in the electromagnetic calorimeter. In general, most energy deposits are there and the cell structure is the same throughout the SPD, PS and electromagnetic calorimeter. After clustering, corrections can be applied by additional algorithms, *e.g.* adding other energy deposits in the cells of the other calorimeter sub-detectors, called digits.

Clusters are searched for by determining local digit maxima in the electromagnetic calorimeter with a transverse energy larger than 50 MeV. These local maxima are used as a starting point. Neighbouring cells with energy deposits are tagged to a local maximum with a Cellular Automaton algorithm to form together a cluster [106]. Basically, the algorithm looks at an untagged cell and tags it using a set of rules based on its relation to neighbouring cells. This is iterated until all untagged cells with non-zero energy deposit are tagged. The definition of what a neighbouring cell is has influence on local maxima determination and subsequent clustering and reconstruction. An example of clustering and the use of different neighbour definitions is illustrated in fig. 8.3. The default is a 3x3 square cluster definition, both for barycentre and energy determination. Since calorimeter showers can overlap, cells are allowed to be assigned to multiple clusters. In this case, the measured energy in the cell has to be divided and therefore each cluster entry has a fractional energy assigned. The fractional energy is determined by using analytical shower shapes (fixed analytical energy distributions) which can be corrected according to the assigned hypothesis of a photon, electron or two photons with overlapping showers coming from a π^0. With this knowledge, the energy

Figure 8.3: Illustration of clustering with different definitions of neighbours for Cellular Automaton clustering algorithm. Numbers indicate amount of energy deposit. Colors denote cells tagged to a specific cluster. Picture from [106].



and the coordinates of the barycentre of the cluster are given by,

$$E = \sum_i f_i E_i, \qquad x = \frac{1}{E}\sum_i x_i \cdot f_i E_i, \qquad y = \frac{1}{E}\sum_i y_i \cdot f_i E_i, \tag{8.5}$$

where E_i, x_i, y_i and f_i denote the measured energy, x- and y-coordinate of the center and assigned fraction of energy of the individual electromagnetic calorimeter cells respectively.

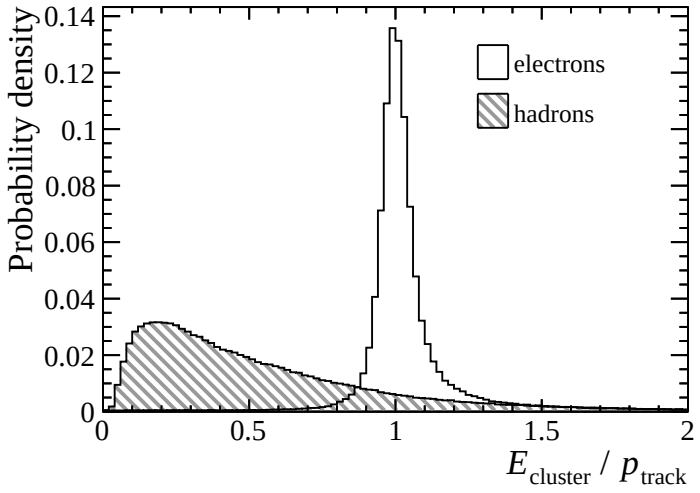
With a cluster, a so called `CaloHypo` object is made to make final corrections to the x-, y- and z-coordinate and the energy of the cluster depending on the hypothesis, *e.g.* photon, electron or positron. These corrections depend on the shower shape fit, possible associated SPD hit and PS energy deposit and the energy of the cluster.

Particle identification

For the purpose of separating electrons from photons, a matching is performed between tracks containing T-station hits and clusters. The x- and y-coordinate and the energy of the cluster, assuming a photon `CaloHypo`, is compared to the extrapolated position and momentum of the track respectively. A matching χ^2, χ^2_{3D}, is constructed by extrapolating tracks towards the z-coordinate of the `CaloHypo`. The covariance matrices, *e.g.* used for the χ^2, are constructed with the size, gain, relative gain and noise per cell for the cluster and the covariance matrix of the track. A correction is applied on the x-coordinate difference based on the momentum and charge of the track to account for track curvature [107].

Apart from discriminating clusters coming from an electron or photon, it also allows to discriminate electrons from hadrons and muons, since electrons loose practically all their energy in the electromagnetic calorimeter, whereas hadrons do not. Hadrons loose also a considerable amount of energy in the hadronic calorimeter as well and muons loose very little in general. Therefore, the variable $E_{\text{cluster}}/p_{\text{track}}$ should peak at one for electrons and have a lower and more widely distributed value for hadrons and muons, as shown in fig. 8.4. Note that this variable is part of the matching χ^2, χ^2_{3D}.

Figure 8.4: Distribution of $E_{\text{cluster}}/p_{\text{track}}$ for electrons (open) and hadrons (hashed) in simulation.



8.3 Bremsstrahlung recovery

Since the calorimeter opens up the possibility to reconstruct photons, radiative losses can be reconstructed. The main goal of bremsstrahlung recovery is to add the missing momentum of these radiative losses to the electron candidate. It consists of two steps. First, it starts by finding electromagnetic calorimeter-clusters that are consistent with coming from the electron. Secondly, the goal is to add the momentum of the matching clusters at the right state of the track, as the loss can in principle happen anywhere along the track.

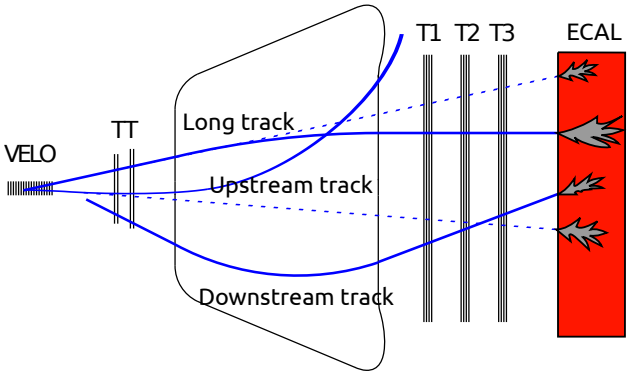
The track before and after the magnet is mostly straight. Therefore, most of the bremsstrahlung photons end up at the intersection of this straight line and the calorimeter. The reason is that there is a relatively low amount of material in the magnet. As the photons emitted after the magnet barely influence the momentum measurement and end up in the area of the cluster of the electron, these photons are not relevant. Hence, the pre-magnet photons coming from the approximately strait line before the magnet most interesting. The geometry of this pre-magnet emission is illustrated in fig. 8.5.

The bremsstrahlung recovery algorithm, `BremAdder`, first extrapolates towards the plane of the electromagnetic calorimeter from two states before the magnet: the state closest to the beam and the state at the end of the TT-stations. This ensures that the full bending range of the track before the magnet is covered. Clusters are then selected that have their center within two standard deviations from the extrapolation area in the xy plane. This standard deviation is determined with the uncertainty on the track parameters determined from the track fit and the cluster spread. From these candidates, only clusters passing a loose requirement on photon ID and having a minimal p_{T} of 75 MeV are selected.

For electrons coming from b-hadron decays, usually only one candidate is added. However, for some tracks there is still enough bending in the TT-stations that can result in separate clusters in the electromagnetic calorimeter. Only about 1% of the times more photons are added.

After the selection of the bremsstrahlung candidates, the momenta of these clusters have to be added to the electron candidate. To determine the momenta of the photons

Figure 8.5: Illustration of the type of electron tracks and associated electromagnetic calorimeter clusters, either from the electron or bremsstrahlung photon (dashed line) in LHCb. Emission of bremsstrahlung is usually before or after the magnet. The latter ends up in the track cluster.



from the energy deposit in the electromagnetic calorimeter, the origin is needed. As default for individual photons, the interaction point is used as origin. If the photon is assumed to come from a decay with additional charged particles, the decay vertex is used as origin. In case of bremsstrahlung, the photon can come from any state of the track. Therefore, the bremsstrahlung photons are matched to the state of the track with the least difference between the slope in the bending plane, t_x, of the track and of the direction of the photon, calculated assuming it comes from the x, y, z position of the state.

Particle identification with bremsstrahlung

Next to momentum reconstruction, bremsstrahlung recovery also serves as information for particle identification. As practically only electrons emit bremsstrahlung, the fact that one has a bremsstrahlung candidate and its corresponding matching χ^2 have discriminatory value, as shown in fig. 8.6. This information is added both in the likelihoods and artificial neural nets used for particle identification. Currently, this matching χ^2 is calculated by linearly extrapolating the track from the state closest to the end of the TT-stations, at z_{endTT}, to the calorimeter plane, $\vec{x}_{\mathrm{track}}$. Subsequently, it is compared to the cluster location, $\vec{x}_{\mathrm{CALO}}$, by minimizing the resulting χ^2 by varying $\vec{x}$,

$$\begin{aligned}\chi^2_{\mathrm{Brem}} = \min\ [&(\vec{x}_{\mathrm{CALO}} - \vec{x})^{\mathrm{T}} C^{-1}_{\mathrm{CALO}} (\vec{x}_{\mathrm{CALO}} - \vec{x}) + \\ &(\vec{x}_{\mathrm{track}}[z_{\mathrm{endTT}}] - \vec{x})^{\mathrm{T}} C^{-1}_{\mathrm{track}} (\vec{x}_{\mathrm{track}}[z_{\mathrm{endTT}}] - \vec{x})]\ ,\end{aligned} \tag{8.6}$$

where C_{CALO} and C_{track} are the covariance matrices of the cluster center assuming it is a photon and the extrapolated track state respectively. However, with current version of the `BremAdder`, more information is known of the matching bremsstrahlung photons. One can calculate a matching χ^2 according to a linear extrapolation from the state to which it matches best, at z_{origin}, between the first state and the state at the end of the TT, by,

$$\chi^2_{\mathrm{BremOrigin}} = (\vec{x}_{\mathrm{CALO}} - \vec{x}_{\mathrm{track}}[z_{\mathrm{origin}}])^{\mathrm{T}} (C_{\mathrm{CALO}} + C_{\mathrm{track}})^{-1} (\vec{x}_{\mathrm{CALO}} - \vec{x}_{\mathrm{track}}[z_{\mathrm{origin}}])\ , \tag{8.7}$$

Figure 8.6: Distribution of bremsstrahlung matching χ^2 for electrons (open) and hadrons (hashed) in simulation. From [108].

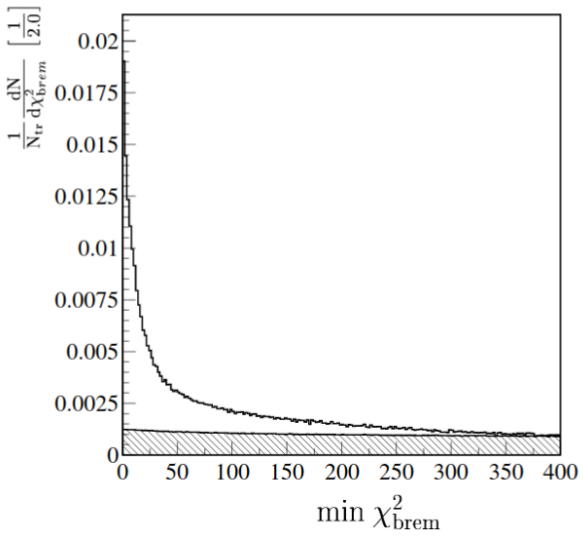
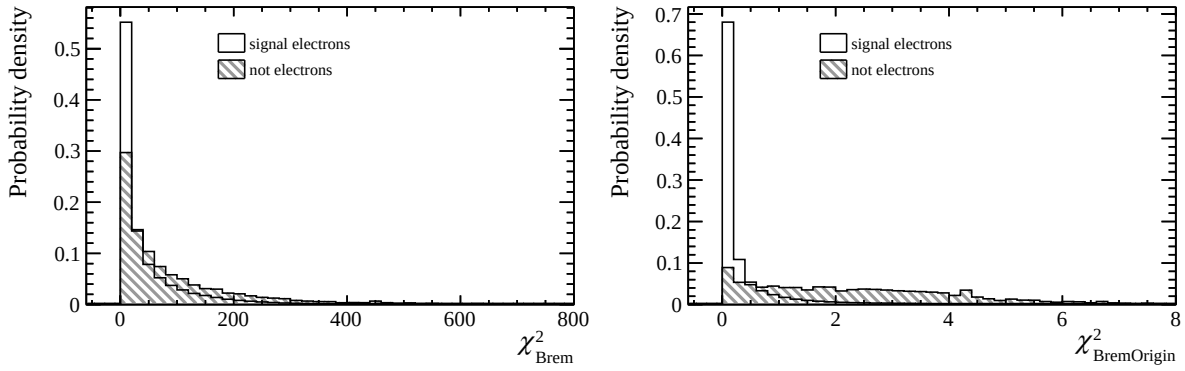


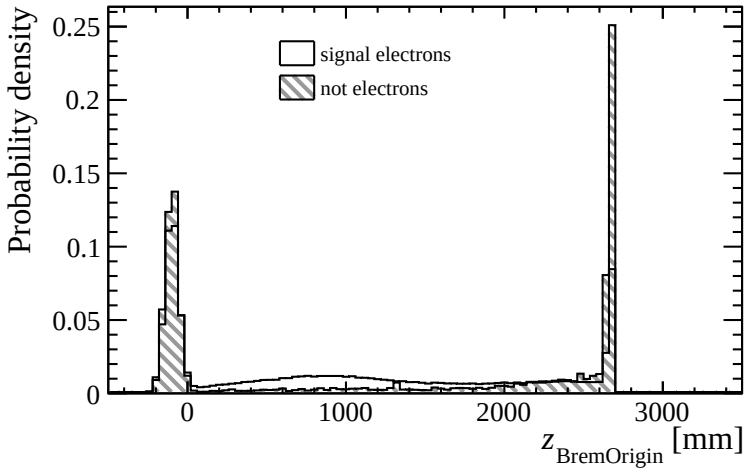
Figure 8.7: Distribution of bremsstrahlung matching χ^2, for electrons (open) and hadrons (hashed) that have a bremsstrahlung match in simulation of $B^0 \to J/\psi(\to e^+e^-)K^{*0}$ decays. On the left, the matching χ^2, χ^2_{Brem}, is shown that is used as input for particle identification. On the right, the updated version of χ^2, $\chi^2_{\text{BremOrigin}}$, is shown that uses the matched track state as used by `BremAdder`. A clear improvement is seen in separation power.



where for C_{CALO} the spread of the cluster is taken. With this redefinition, a better performance is seen, as can be seen in fig. 8.7.

In addition, also the location of the matched state has discriminatory value, as seen in fig. 8.8. This is caused by that fact that bremsstrahlung more likely comes from a z-coordinate where there is more material, while fake bremsstrahlung is more likely to come from a z-coordinate where there is more bending in the track. The latter is at higher z, as it gives a wider range in the calorimeter plane per step in z, which in turns give a higher probability to match to a random cluster.

Figure 8.8: Distribution of the z-coordinate of the state to which the bremsstrahlung photon candidate matches best for electrons (open) and hadrons (hashed) that have a bremsstrahlung match in simulation of $B^0 \to J/\psi(\to e^+e^-)K^{*0}$ decays.



8.4 Reconstructed electrons

Practically, electrons used in analyses at LHCb come in three forms of track reconstruction: long, downstream and upstream. These cover electrons from the following sources: from a b-hadron or c-hadron decay, or a photon conversion. Other reconstructable object can be associated to tracks like calorimeter clusters from either the track or bremsstrahlung, as discussed previously in section 8.2, but also particle-identification likelihood-ratios obtained from hits from Cherenkov photons in the RICH1 or RICH2, or muon station track segments, *i.e.* the absence of.

In most cases, electrons come in the form of long tracks, which can have all possible particle-identification methods. Downstream tracks, primarily for photon conversions, also have all possible particle-identification methods available. This is the case because, despite the RICH1 not being within the direct range of the track, the track can still be extrapolated upstream to the RICH1 and hits can be associated. In case of upstream tracks, they usually have low momentum and bend out of the acceptance of the T-stations. But they also consist of tracks that pass through the T-stations but do not pass the forward or matching pattern-recognition. These tracks don't have calorimeter cluster, RICH2 or muon station information associated to it. For electrons, upstream tracks are of special interest and they will be discussed in section 8.4.1. The track types with their associable particle-identification objects are listed in table 8.1.

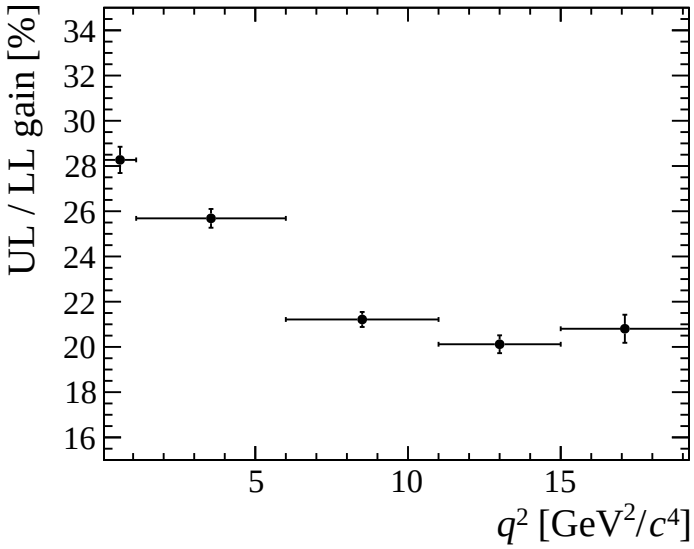
8.4.1 Upstream

While in general low-momentum tracks are not of much interest, due to bremsstrahlung recovery, the momentum of electron tracks determined with the track fit might be low, while the fully reconstructable momentum might not be. In the next chapter, chapter 9, efficiency loss of long reconstruction is discussed. This efficiency loss is partially recoverable by using upstream tracks, as low momentum tracks might bend out of the detector after the magnet. Nonetheless, before the magnet at the TT-stations, they are less likely to be

Table 8.1: Types of tracks and their available particle identification options (for electrons).

Track Type	Particle identification				
	RICH1	RICH2	Calorimeter		Muon
			Track cluster	Bremsstrahlung	
Long	✓	✓	✓	✓	✓
Downstream	✓	✓	✓	✓	✓
Upstream	✓			✓	

Figure 8.9: Gain of using upstream tracks, long plus upstream for the e^+e^- combination, in $B^0 \to K^* e^+ e^-$ decays depending q^2, *i.e.* the invariant mass squared of the e^+e^- combination.



out of acceptance. Such an efficiency gain, *i.e.* extra amount over original amount, leads to a double efficiency gain for an e^+e^- combination. A good example of where efficiency can be gained is for the analysis of $R_{K^{*0}}$. The gain of using an upstream plus long track with respect to using two long tracks for the reconstruction of e^+e^- in $B^0 \to K^* e^+ e^-$ is shown in fig. 8.9, depending on the invariant mass squared, q^2, of the e^+e^- combination. This shows that, especially for the interesting low-q^2 region, significant gains of around 20 to 30% can be obtained.

An advantage of these tracks having low momenta in the TT-stations is that their momentum resolution is also better at low momentum. The absolute pull of the reconstructed momentum and the track momentum distribution, both for long and upstream, is shown in fig. 8.10. Eventually, the measure or variable that matters in regards of the momentum resolution is the resulting resolution of the invariant mass. This is usually used in a maximum likelihood fit to discriminate between signal and background. For $B^0 \to J/\psi (\to e^+e^-) K^{*0}$ decays, the reconstructed invariant mass is shown in fig. 8.11. It shows that indeed the resolution is worse compared to long tracks, but since it is a b-hadron decay, extra information can be obtained by constraining the b-hadron momentum to point to the PV because of its fairly long life time. This is especially helpful for the case

Figure 8.10: (Left) Distribution of the track momentum for long and upstream electron tracks. (Right) Absolute pull of the reconstructed momentum, *i.e.* track fit plus bremsstrahlung recovery momentum with respect to true momentum, for track types long and upstream for electrons versus the measured track momentum. For both figures, a sample of electrons from simulation of $B^0 \to J/\psi(\to e^+e^-)K^{*0}$ decays is used.

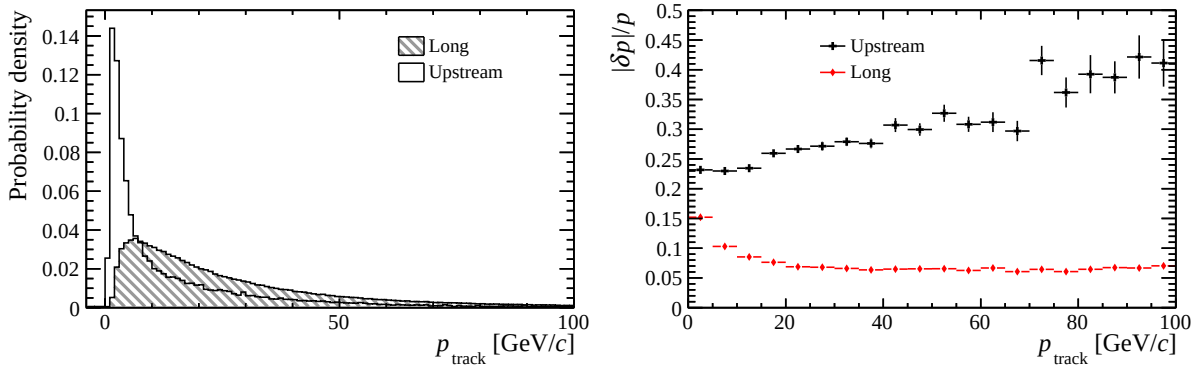
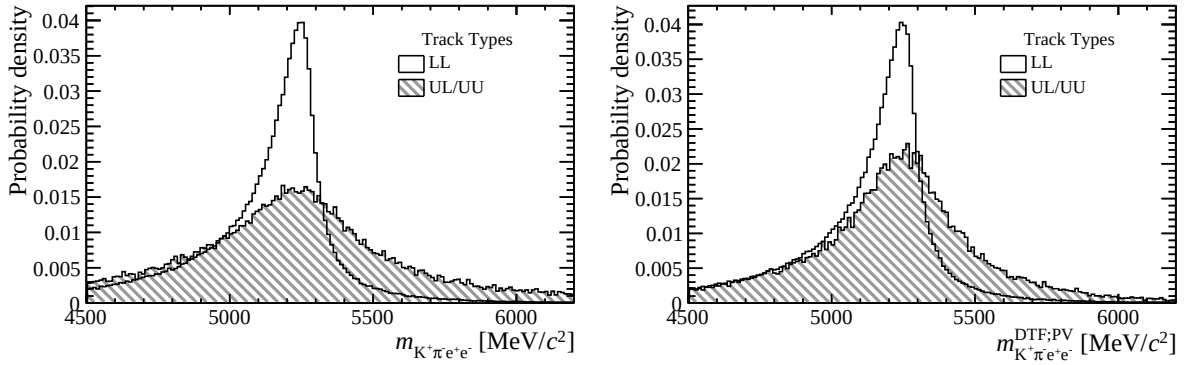


Figure 8.11: Invariant mass distributions of $B^0 \to J/\psi(\to e^+e^-)K^{*0}$ decays in simulation with two long electron tracks (open) and for at least one upstream electron track (hashed). With (left) and without (right) constraint to PV using `DecayTreeFitter` [64].

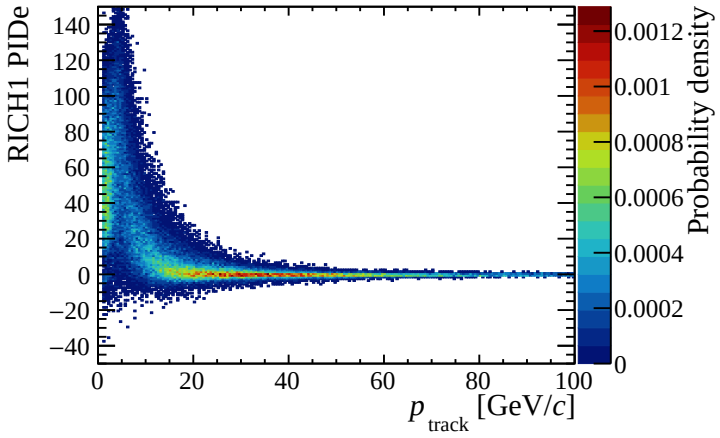


of upstream electrons. An important notion is that the low-mass tail, where usually most backgrounds are, is not larger compared to the invariant mass using only long tracks.

Aside from reconstructing the momentum, the electron candidate needs particle identification. As discussed earlier, upstream tracks have bremsstrahlung and RICH1 information available. The improvement in particle identification with bremsstrahlung photons, as discussed previously in section 8.2, is of particular importance for the upstream electrons tracks. Next to this, the RICH1 information is more discriminatory at low momentum, as can be seen from fig. 8.12. Here, the logarithm of the electron-identification likelihood-ratio, with respect to a pion is shown, depending on the track momentum. From this, one can see that below a momentum of 10 GeV/c, the discriminatory value becomes quite significant. The majority of the upstream tracks is below this value.

On the whole, the use of upstream tracks for electron reconstruction shows a promising way to increase the yield of decays containing electrons at LHCb, especially for rare decays like $B^0 \to K^* e^+ e^-$.

Figure 8.12: Logarithm of the electron identification likelihood ratio, PIDe, from RICH1 of upstream electron tracks in simulation of $B^0 \to J/\psi(\to e^+e^-)K^{*0}$.



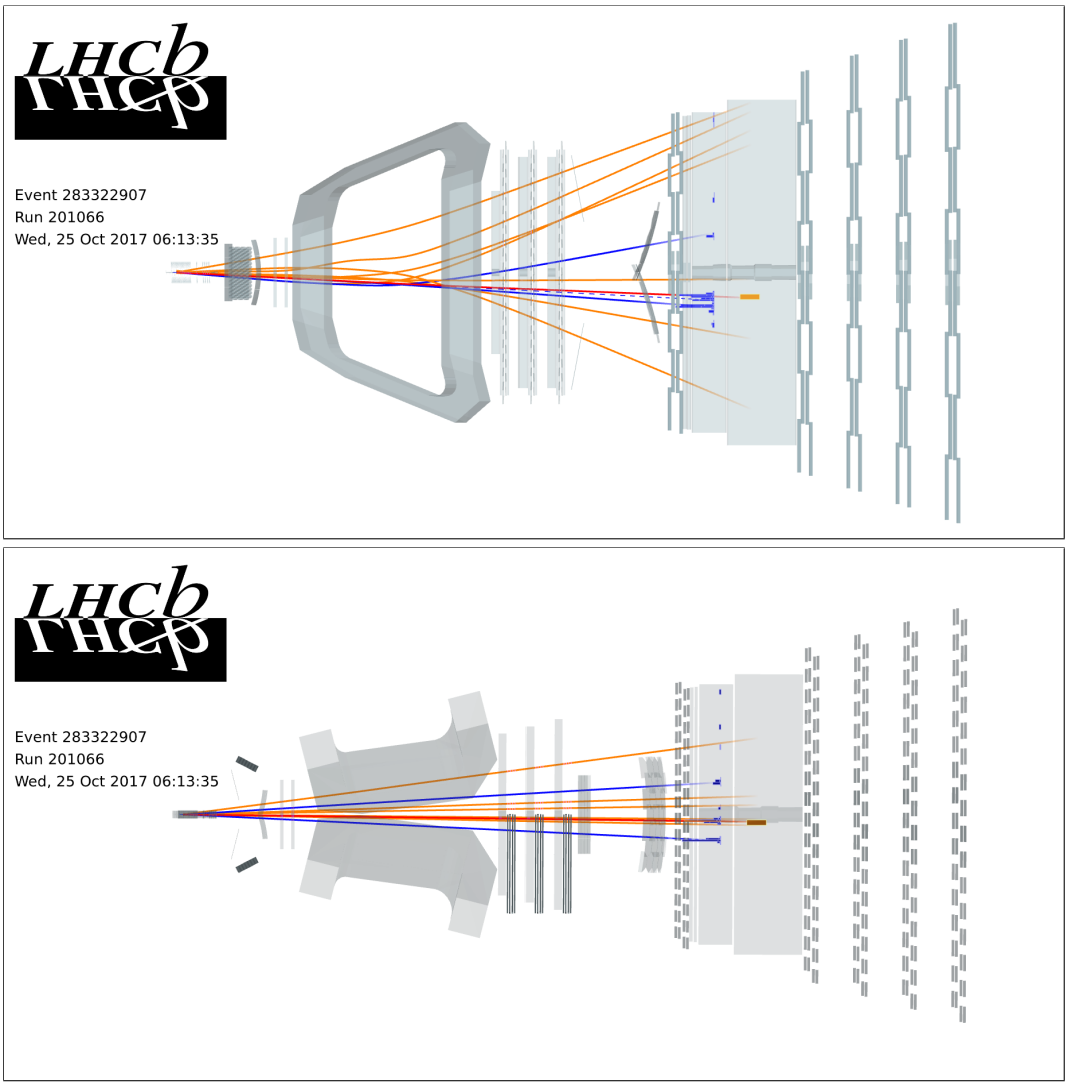
8.5 Visualisation: *event display*

At this point, the reconstruction relevant for electrons at LHCb has been described. To illustrate how the reconstructed objects relate to raw data and the detector in general, an event which has a high probability of containing a $B^+ \to J/\psi(\to e^+e^-)K^+$ decay in 2017 data is shown with the LHCb event display in figs. 8.13, 8.14 and 8.16. In addition, the occupancy of the event is chosen to be low, to clearly see the signal particles and their associated reconstructed objects and raw data. This event has one PV, 10 long tracks, 5 downstream tracks and 86 hits in the SPD.

In the first set of illustrations, fig. 8.13, from top to bottom, the top view and side view are shown. For the top view and side view, one can see the long tracks as extrapolations from, in z, the PV, to before the calorimeter using splines. The color of the tracks are determined according to their most likely hypothesis determined with PID variables with blue for electrons, red for kaons, orange for pions, purple for protons and green for muons. Next to the charged tracks, bremsstrahlung photon candidates are shown with blue dashed lines, extrapolated from the start of the associated electron track candidate to the calorimeter. In addition, energy deposits in the electromagnetic calorimeter and the hadronic calorimeter are shown in blue and orange respectively. The bar scales with the energy of the deposit. Hits in the muon stations are shown in green, although there are no muon hits in this particular event. In this case, especially in the zoomed view of the calorimeter in fig. 8.16, one can clearly see one of the electron candidates with a high-energy, 59 GeV, track-associated electromagnetic calorimeter cluster and 24 GeV bremsstrahlung cluster. The other electron is also visible with clearly a lower energy. The kaon candidate also has an electromagnetic calorimeter cluster associated to it, but also clearly has a hadronic calorimeter energy deposit behind it.

To get a better idea of how the displaced nature of the b-hadrons is seen in the detector, a zoom of the VELO is shown in figs. 8.14 and 8.15, with an additional zoom into the PV and secondary vertex area. The top illustration is the top view and the lower one is the side view. For the zoomed in illustration, a small circle is used for the secondary vertex of the $B^+ \to J/\psi(\to e^+e^-)K^+$ candidate and a larger circle for the PV. The extrapolation from the PV clearly shows the impact parameter of the tracks that come from a secondary

Figure 8.13: Event display, containing a high-probability $B^+ \to J/\psi(\to e^+e^-)K^+$ candidate, with view of the full detector from the top (top) and the side (middle).



vertex. Determined with the LHCb reconstruction, some characteristics of the signal candidate in the event are calculated and listed in table 8.2.

To get a better impression of an event with a more average occupancy, another event is shown in fig. 8.17. This event has one PV, 40 long tracks, 11 downstream tracks, 2 muon tracks, and 377 hits in the SPD.

Figure 8.14: Event display, concentrated on the VELO, with zoom where the PV and the secondary decay vertex are highlighted with a big and small circle respectively. Event is the same as fig. 8.13.

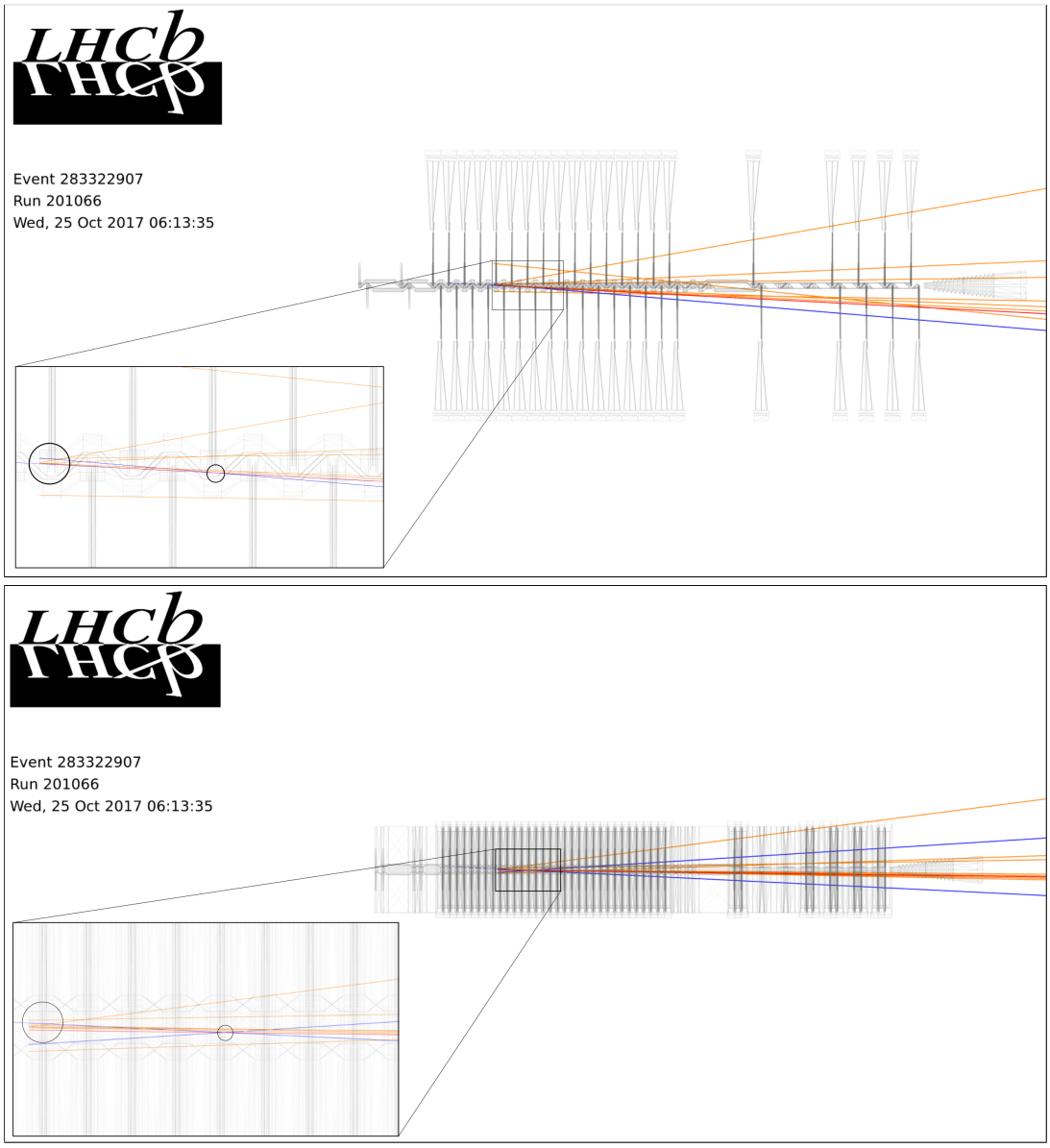


Figure 8.15: Event display, with zoomed view of the VELO in the xy plane. The PV and the secondary decay vertex are highlighted with a big and small circle respectively. Event is the same as fig. 8.13

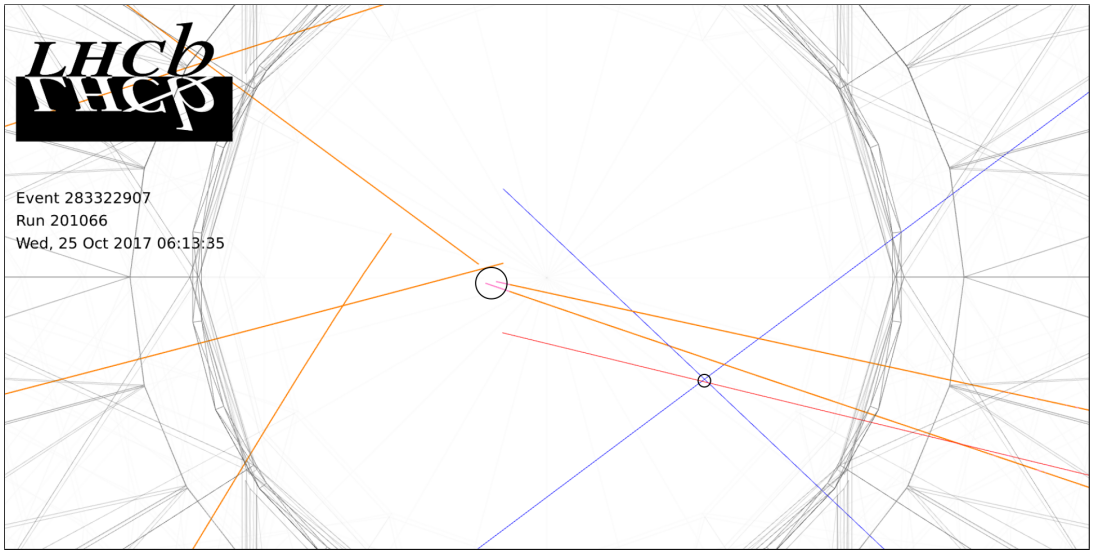


Figure 8.16: Event display, with zoomed view of the calorimeter region. Event is the same as fig. 8.13

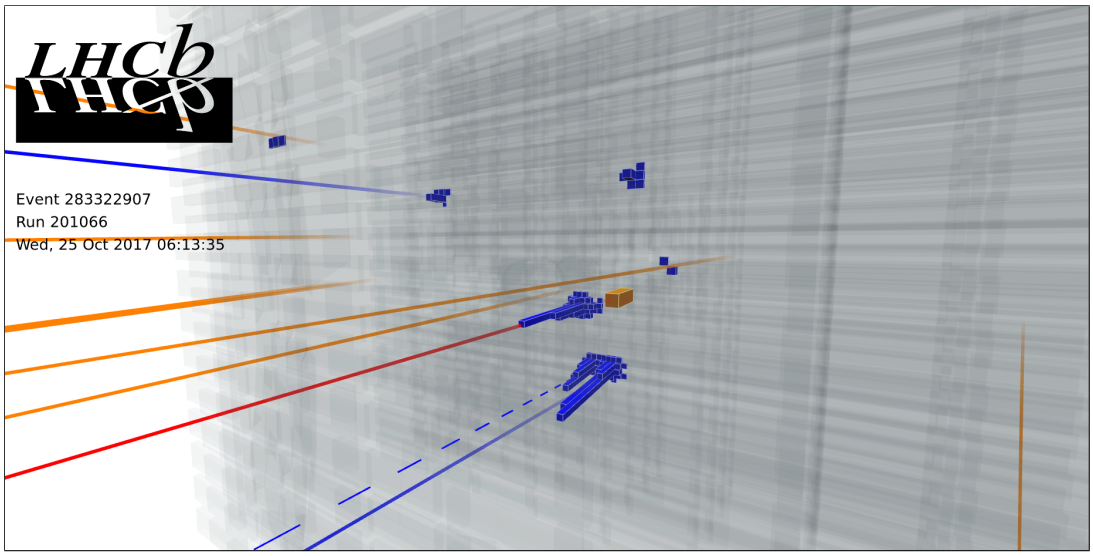
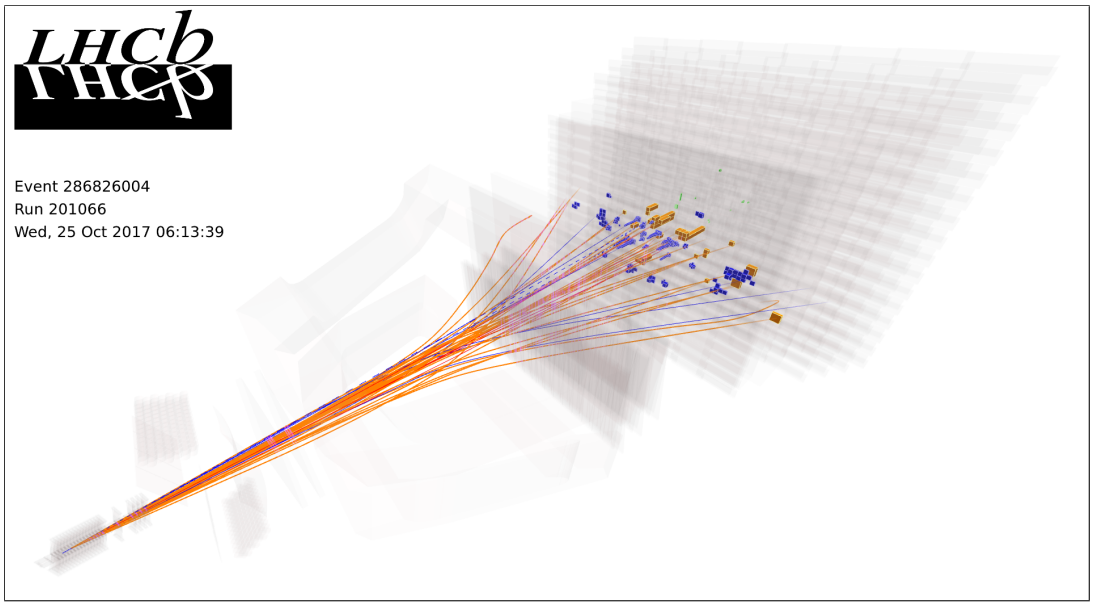


Table 8.2: Some of the variables from reconstruction of the $B^+ \to J/\psi(\to e^+e^-)K^+$ candidate in event 283322907 of run 201066 displayed in figs. 8.13, 8.14 and 8.16. A part of the variables are related to chapter 9.

Observable	Value	Unit	Notes
		probe (VELO track)	
IP	6.059	mm	
χ^2_{IP}	6204.23		
$\chi^2_{\text{track}}/n_{\text{DOF}}$	0.16		
		long track match to probe	
nMatch/nID	100	%	percentage of LHCbID overlap in VELO between probe and long track
p_{track}	4.819	GeV/c	
$\chi^2_{\text{track}}/n_{\text{DOF}}$	0.53		
ProbNNe	0.9999		
$E_{\text{cluster}}/p_{\text{track}}$	1.06		
		tag electron	
IP	1.598	mm	
χ^2_{IP}	2843.69		
$\chi^2_{\text{track}}/n_{\text{DOF}}$	1.14		
p_{track}	58.533	GeV/c	
p_{brem}	24.274	GeV/c	
ProbNNe	0.206		
		K^+	
IP	0.829	mm	
χ^2_{IP}	778.41		
$\chi^2_{\text{track}}/n_{\text{DOF}}$	0.83		
p	107.772	GeV/c	
ProbNNk	0.5054		
		B^+	
$\chi^2_{\text{vtx}}/n_{\text{DOF}}$	2.19		
χ^2_{IP}	0.31		
FD	66.716	mm	flight distance
$m^{DTF}_{e^+e^-}$	2970	MeV/c^2	constrained to PV
$m^{DTF}_{e^+e^-}$	3081	MeV/c^2	constrained to B^+ mass
$m^{DTF}_{J/\psi K^+}$	5283	MeV/c^2	constrained to PV and J/ψ mass
$p^{\text{DTF}}_{\text{probe}}$	8.519	GeV/c	constrained to PV and J/ψ mass

Figure 8.17: Event display of an event in data containing a $B^+ \to J/\psi(\to e^+e^-)K^+$ candidate where the event has a more regular occupancy. The event has one PV, 40 long tracks, 11 downstream tracks, 2 muon tracks, and 377 hits in the SPD.



Chapter 9

Reconstruction efficiencies

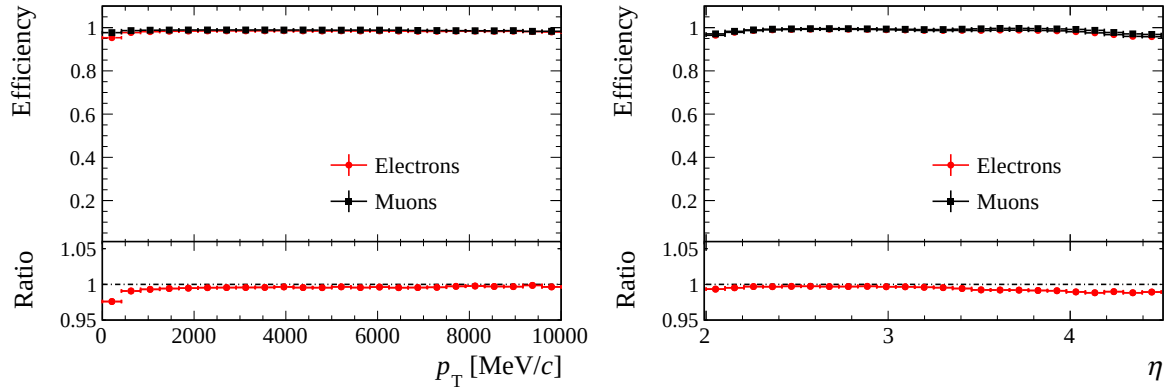
In the past at LHCb, the efficiency of reconstructing a stable charged particle has been taken from either simulation or simulation corrected with tracking efficiencies obtained with a data-driven method that only covers muons [109]. In recent years, an effort has been made to introduce new data-driven methods that cover not only tracking efficiencies, but also the more inclusive reconstruction efficiency of not only muons, but hadrons and now also electrons.

The method to determine reconstruction efficiencies for electrons, which has been published in [110], will be discussed in this chapter. It covers electrons that are produced within the VELO, covering the ones originating from pp collisions and from consecutive b-hadron and c-hadron decays. Hence it covers the majority of analyses using electrons at LHCb.

The focus of the method is to measure the ratio of efficiencies between data and simulation and not the absolute efficiencies, since for the ratio systematic uncertainties are lower. With these efficiency ratios, corrections can be applied to simulation and systematic uncertainties can be evaluated.

In section 9.1, the efficiencies of different track types in simulation are introduced to assess and define what efficiency is needed to be determined. Second, the method (tag-and-probe with VELO tracks) will be introduced in section 9.2. To use these efficiencies, *e.g.* to correct simulation, it is needed to unfold these efficiencies in terms of the kinematics of the electron at production. The inference of the kinematics of the probe, especially its absolute momentum, is discussed in section 9.3. The selection of probe electrons for the implementation of the method is defined and explained in section 9.4 and the method to extract the efficiencies from the data is described in section 9.5. The definition of the efficiency combined with the behaviour of electrons has a subtlety due to contributions from fake tracks (ghosts). The origin, consequences, and treatment of ghosts, both in this analysis and others, are discussed in section 9.6. The results of the method, in the form of efficiencies from data and simulation in terms of kinematics, will be presented in section 9.7. The validation of the method, the determination of systematic uncertainties and cross-checks associated to the results will be discussed in section 9.8. A use case and further possibilities will be discussed in section 9.9, finishing with a conclusion.

Figure 9.1: Efficiencies in simulation to reconstruct leptonic VELO tracks, assuming the lepton has $\eta \in [1.9, 4.5]$ for both figures. Samples are simulated $B^+ \to J/\psi(\to e^+e^-)K^+$ and $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ decays for the electron and muon respectively. The ratio is the efficiency of reconstructing electrons over the one for muons.



9.1 Efficiencies in simulation

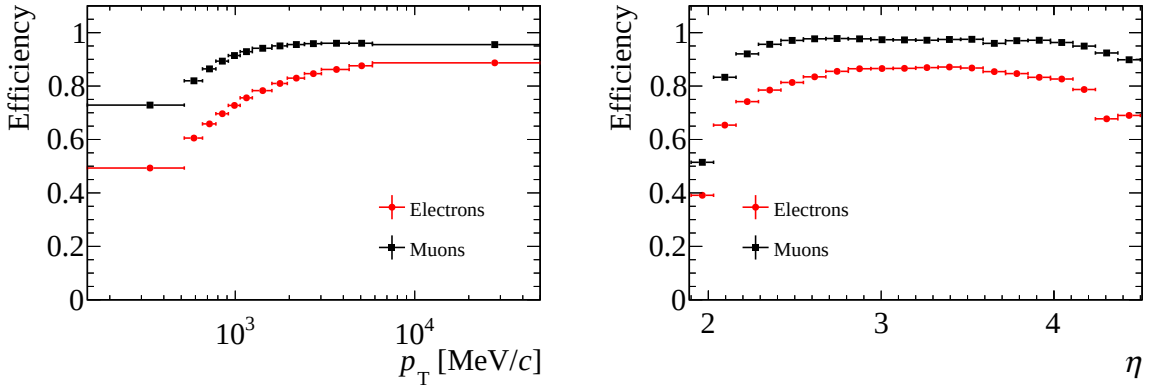
Electrons originating from b-hadron and c-hadron decays (that are within the geometrical acceptance of LHCb) travel first through the VELO. There, their tracks have a high efficiency, fairly independent of kinematics, as can be seen in fig. 9.1. Furthermore, the efficiency is not much different from that of muons. The difference stays within 1%. Specifically for VELO tracks, a determination of the efficiencies in data has been performed, using muons [111]. This determination for muons can be extrapolated to electrons using an uncertainty on the relatively well known material budget of the VELO of about 6%, leading to a systematic uncertainty of 0.06% on the extrapolated efficiency.

It is after the VELO, where charged particles are bend by the magnet, that the momentum of the particle and subsequently its momentum loss along its trajectory become more important. Accordingly, the efficiency to reconstruct electrons as long tracks is much more affected than *e.g.* muons, as can be seen from fig. 9.2. Hence, it shows the need to have a dedicated efficiency measurement for electrons, compared to muons of which there are already data-driven methods to measure long tracking efficiencies [109]. Note that a tracking efficiency is not the same as a reconstruction efficiency, as a tracking efficiency is conditional on that the track should have travelled through the required tracking stations. Therefore, it shows the efficiency of tracking stations and subsequent track finding algorithms, whereas a reconstruction efficiency also includes the case where tracks travel through non-instrumented areas.

The long-track reconstruction efficiency shows a dependence both on pseudo-rapidity, η, and transverse momentum, p_{T}. As their distributions can differ significantly between decay channels, the unfolding of the dependence of the efficiency on η and p_{T} is necessary. The distribution of the remaining degree of freedom, the azimuthal angle ϕ, is also affected. Differences between decay channels in terms of ϕ are negligible and its distribution is fairly flat. Despite that, a dependence on ϕ is considered in this analysis, as the amount of material does have a ϕ dependence.

To conclude, the goal of this analysis is to measure the efficiency to reconstruct an electron as a long track, given that it is reconstructed as a VELO track, hereafter referred

Figure 9.2: Efficiencies in simulation to reconstruct long tracks, assuming lepton has $\eta \in (1.9, 4.5)$ for both figures. Samples are $B^+ \to J/\psi(\to e^+e^-)K^+$ and $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ decays simulated with ReDecay [112].



to as the electron reconstruction efficiency.

9.2 Tag-and-probe method with VELO tracks

To be able to measure the electron reconstruction efficiency, it is essential to get an unbiased electron VELO track, while at the same time one has to be able to distinguish the track from background. To do this, the tag-and-probe method is used. Essentially, a decay is chosen which contains the probe particle, this case an electron, but the fact that it is an electron can be deduced by the rest of the particles in the decay, called the tag.

It is ideal to tag an electron in a decay of a neutral resonance to e^+e^-. So far no charged lepton number violation is seen. This ensures that if one has an electron with an additional track that comes from the neutral resonance, one knows that the additional track is also an electron. A very common neutral resonance that has a high branching fraction to e^+e^- is the J/ψ. Decays of b-hadrons contain a J/ψ about 1% of the time and the branching fraction of $J/\psi \to e^+e^-$ is about 6% [16]. A big advantage of a J/ψ coming from a b-hadron is that, due to the lifetime of b-hadrons, the final state particles form a displaced vertex. This makes it geometrically distinguishable from backgrounds. In addition, the tag electron can be used at `L0` level. This sets indirectly a particle-identification requirement on the electron, as muons are highly unlikely to fire this trigger. For this analysis the $B^+ \to J/\psi(\to e^+e^-)K^+$ decay is used. The application of the method on this decay is illustrated in fig. 9.3. The selection is further described in section 9.4.

Despite a loose selection on the probe track, it is still correlated to the tag particles through the decay kinematics. This correlation is covered by unfolding the efficiency in terms of kinematics, which is already necessary for the use of this analysis for efficiency corrections for other analyses in the first place. The unfolding is performed by binning in the phase space of the probe electron. To determine the binning, two points are crucial. First, it is needed to cover the variations of the efficiency. Subsequently, the efficiency should not vary significantly compared to the uncertainty on the efficiency. Second, the in-bin differences of the kinematics of simulation versus data need to be small enough to be able to compare the efficiencies. To cover these criteria, it is chosen that the variation of

Figure 9.3: Illustration of the tag-and-probe method. Signature is a decay, this case $B^+ \to J/\psi(\to e^+e^-)K^+$, with tag particles that ensure that the probe is an electron. The selection is based on the displaced nature of the vertex formed by the tag particles. The trigger makes use of energy deposits of the tag electron in the electromagnetic calorimeter.

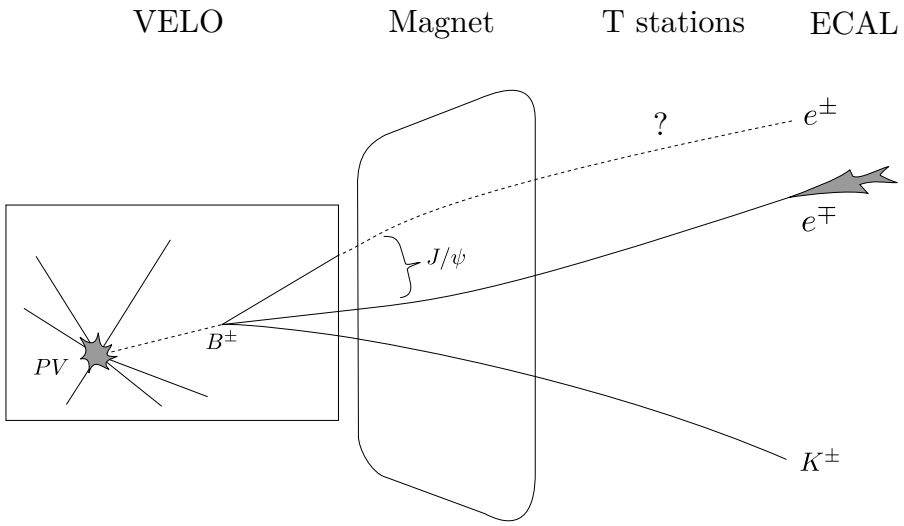


Table 9.1: Binning in p_{T} and η. The bins in p_{T} are optimised per η bin based on the efficiency variation.

η	p_{T} bin edges [GeV/c]
$1.9 \leq \eta \leq 2.9$	$0.15, 0.52, 0.657, 0.780, 0.915, 1.065,$
	$1.245, 1.6, 1.95, 2.43, 3.02, 4.31, 5.81, 50$
$2.9 \leq \eta \leq 3.5$	$0.15, 0.865, 1.475, 4, 50$
$3.5 \leq \eta \leq 4.0$	$0.15, 0.49, 0.760, 1.3, 2, 50$
$4.0 \leq \eta \leq 4.5$	$0.15, 0.55, 0.785, 50$

the efficiency in simulation (first half versus second half of the bin) does not exceed 3% or that the variation is larger than five times the statistical uncertainty. In case the variation is larger, the bin is split at the median of the p_{T} distribution in the bin. To ensure the bin has enough signal to apply a likelihood fit, a minimum on the signal yield is required. This optimisation is applied per bin in η. For ϕ, which is practically uncorrelated to p_{T} and η, two bins are chosen: one bin that contains more material due to the RF-foil, the foil separating the VELO vacuum from the beam vacuum, and the other bin containing the rest. The binning in p_{T} and η is shown in fig. 9.4 and table 9.1. For ϕ the binning is listed in table 9.2.

Figure 9.4: Efficiencies in optimised binning for $1.9 \leq \eta \leq 2.9$ (top left), $2.9 \leq \eta \leq 3.5$ (top right), $3.5 \leq \eta \leq 4.0$ (bottom left) and $4.0 \leq \eta \leq 4.5$ (bottom right).

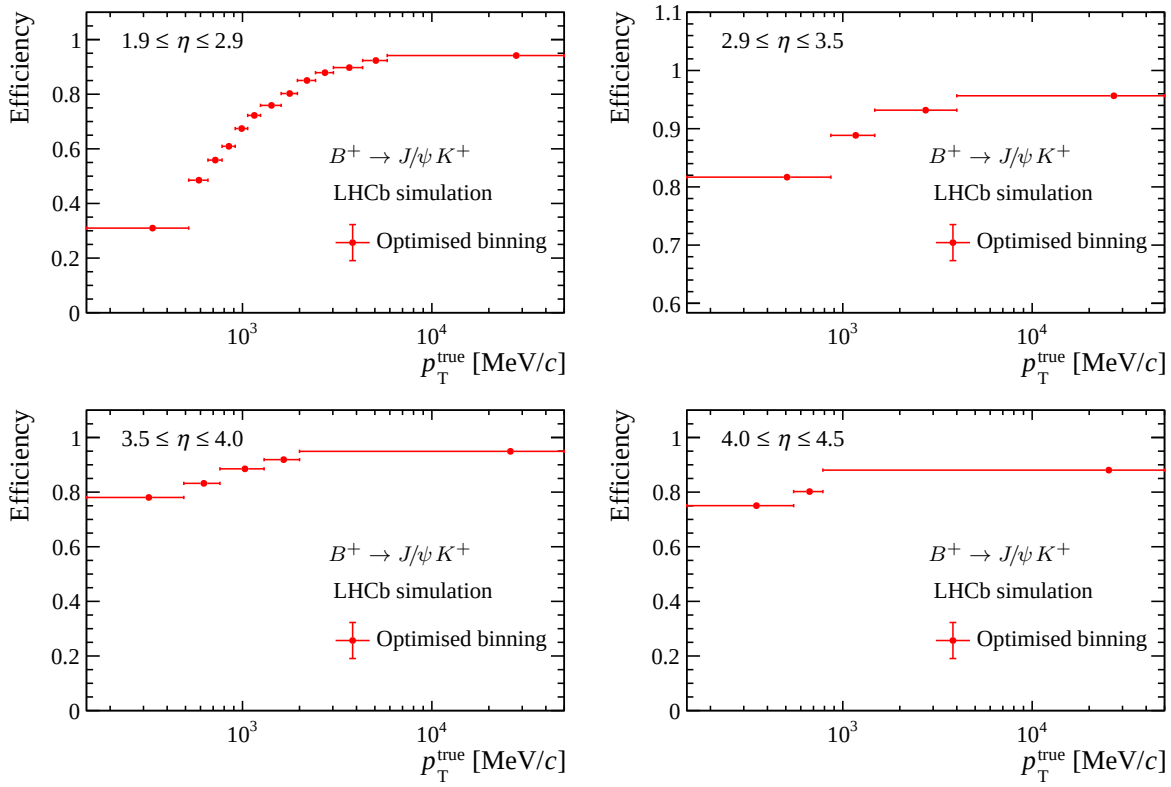


Table 9.2: Binning in ϕ. Two regions are chosen based on their material content.

Name	Bin definition
RF-foil	$\lvert\phi - \pi/2\rvert < \pi/8$ `or` $\lvert\phi + \pi/2\rvert < \pi/8$
Non-RF-foil	$\lvert\phi - \pi/2\rvert \geq \pi/8$ `and` $\lvert\phi + \pi/2\rvert \geq \pi/8$

9.3 Momentum inference

Determining the momentum of the probe has a two-fold importance. First of all, it is needed to determine the efficiency dependence on kinematics. Second, the momentum is needed to get an invariant mass for the likelihood fit to obtain the efficiency.

It is essential that the momentum that is to be inferred is the momentum of the electron at production, such that the kinematic dependence of the efficiency can be unfolded in a way that is independent of the efficiency itself, *i.e.* the momentum before going through the detector. Another criterium is that the momentum at production is independent of the decay. The latter criterium has a subtlety due to final-state radiation, since in data the distinction between bremsstrahlung from material before the magnet, directly correlated to efficiency losses, is very similar to final-state radiation, which is decay specific. Therefore, to determine its resolution, the inferred momentum will be compared to the momentum at decay that does not include final-state radiation.

Since the direction, η and ϕ, of the probe electron is known to negligible precision

from the VELO track fit, only the modulus of the momentum is unknown. Hence, only one constraint is needed. For this purpose, the invariant mass of the dielectron can be constrained to the mass of the J/ψ with the following equation,

$$\begin{aligned} m^2_{J/\psi} &= \left(p^\mu_{\text{probe}} + p^\mu_{\text{tag}}\right)^2 \\ &= \left(\sqrt{m_e^2 + p^2_{\text{probe}}} + \sqrt{m_e^2 + p^2_{\text{tag}}}\right)^2 - p^2_{\text{tag}} - p^2_{\text{probe}} - 2p_{\text{tag}}p_{\text{probe}}\cos\theta \end{aligned} \tag{9.1}$$

where, respectively, p^μ_{probe} and p^μ_{tag} are the four-momenta of the probe and tag electron, $m_{J/\psi}$ and m_e are the mass of the J/ψ and the electron, p_{probe} and p_{tag} are the absolute momenta of the probe and tag and θ is the opening angle between the probe and tag electron. Since $m_e \ll p_{\text{probe}}$, the approximation that $m_e^2/p^2_{\text{probe}} = 0$ can be made with a precision well below other effects on the resolution. This way, the absolute momentum of the probe can be solved and is given by,

$$p_{\text{probe}} \simeq \frac{m^2_{J/\psi} - 2m_e^2}{2\left(\sqrt{m_e^2 + p^2_{\text{tag}}} - p_{\text{tag}}\cos\theta\right)} \,. \tag{9.2}$$

It is used for reasons of speed and simplicity in the trigger, which will be discussed in section 9.4.1. The disadvantage of using this constraint is its dependence on the tag electron. Due to the tag's momentum loss, eq. (9.2) is biased towards higher momenta for the probe. The advantage of just using a J/ψ-mass constraint is that the invariant mass of the full $J/\psi\, K^\pm$ combination is still a degree of freedom. Subsequently, it is usable to distinguish signal from background. For this reason, a J/ψ-mass constraint is used for the invariant mass used for the likelihood fit of the efficiency extraction, described later in section 9.5.

For the determination of the probe momentum for the kinematic dependence of the efficiency, this concern is not relevant, so more constraints can be used. Here, the main concern is to have an unbiased and precise as possible momentum determination. To avoid the bias from the momentum of the tag electron, its momentum is assumed to be unknown and also determined from the constraints. With two degrees of freedom, another constraint is needed, the B^+-mass constraint, which is constructed in an equivalent way to eq. (9.1) by,

$$m^2_{B^+} = \left(\sqrt{m^2_{J/\psi} + p^2_{J/\psi}} + \sqrt{m_K^2 + p_K^2}\right)^2 - p^2_{J/\psi} - p^2_K - 2p_{J/\psi}p_K\cos\theta_K \tag{9.3}$$

where, respectively, $p^\mu_{J/\psi}$ and p^μ_K are the four-momenta of the J/ψ and the kaon, m_K is the mass the kaon, $p_{J/\psi}$ and p_K are the absolute momenta of the J/ψ and the kaon and θ_K is the opening angle between the J/ψ and the kaon. Note that $p^\mu_{J/\psi}$ and consequently $p_{J/\psi}$ are dependent on p_{probe} and p_{tag}.

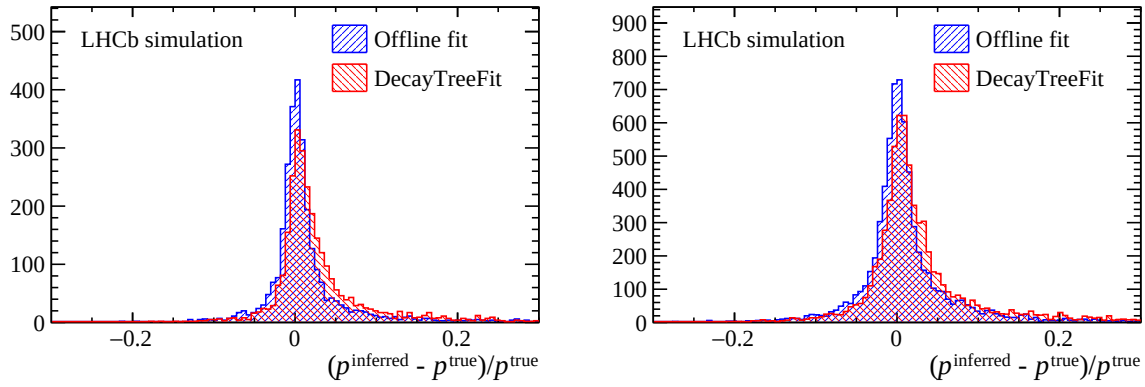
Implementing the constraints of eqs. (9.1) and (9.3) can be done in multiple ways. With LHCb software, it is possible to do a χ^2 fit to the topology of the decay chain, called `DecayTreeFitter`. It can implement constraints to vertex, of both decay and origin, and invariant mass by χ^2 minimisation with a Kalman filter [64]. The `DecayTreeFitter` uses the covariance matrices of the momenta and vertices of the particles that make up the decay. As a χ^2 fit assumes Gaussian uncertainties, this method is biased for electrons. Thus, a more simple application of the mass constraints is implemented. The

probe electron momentum and (as nuisance parameter) the tag electron momentum are determined for the kinematic dependence of the efficiency by using eqs. (9.1) and (9.3) to construct a function, $f(p_{\text{probe}}, p_{\text{tag}})$, that should be minimal. It is given by,

$$f(p_{\text{probe}}, p_{\text{tag}}) = \sqrt{\left(m^2_{B^+} - (p^\mu_{\text{probe}} + p^\mu_{\text{tag}} + p^\mu_K)^2\right)^2 + \left(m^2_{J/\psi} - (p^\mu_{\text{probe}} + p^\mu_{\text{tag}})^2\right)^2} \,. \quad (9.4)$$

and is subsequently minimised using the Nelder-Mead method [113]. This determination is independent of the estimated uncertainties of the probe and tag electron. The method is hereafter called offline fitter. The performance, *i.e.* resolution, of the momentum inference is shown in fig. 9.5, both for the `DecayTreeFitter` and the offline fitter. While the offline fitter reduces the bias from the momentum loss of the tag electron, it should be noted that the irreducible bias from final-state radiation is still there.

Figure 9.5: Resolution of the momentum inference with `DecayTreeFitter` and offline fitter for probe electrons with $p < 3\,\text{GeV}/c$ (left) and $p > 7\,\text{GeV}/c$ (right). The tail to the right side is mostly due to overestimation in case of final-state radiation in $J/\psi \to e^+e^-$.



9.4 Selection of $B^+ \to J/\psi\,(\to e^+e^-)K^+$ decays

The selection used for the method consists of two stages. First of all, an online selection in the trigger has to be applied, which is discussed in section 9.4.1. Further offline selection is applied to increase purity, including some selection directly related to the probe. This is described in section 9.4.2. The definition of what constitutes a pass, *i.e.* that a VELO probe track has an associated long-track, is given in section 9.4.3.

The resulting dataset that is used in this analysis corresponds to $1.3\,\text{fb}^{-1}$ of pp collisions at $\sqrt{s} = 13\,\text{TeV}$ taken in late 2017 with the LHCb detector at the LHC.

9.4.1 Trigger: *TurboCalib*

The effort to start measuring electron reconstruction efficiencies was started in 2017, in the middle of Run 2 of the LHC. The option was either to use the existing dataset and consequently use a trigger selection that is already available, or develop a new selection that was to be implemented in the trigger running at the time.

The latter option was chosen to have the most freedom in developing the method and make a selection that leaves the probe as unbiased as possible, while keeping enough efficiency and have a high enough purity. This option has another more technical advantage because the existing full dataset is not directly accessible. It has to go through so called Stripping campaigns. This essentially means that another step in the data flow has to be done that has to be processed centrally, *i.e.* for the whole LHCb collaboration. This is not repeated often as these campaigns cost many CPU hours and take weeks to complete.

In case of the new trigger line option, this data does not have to go through these steps. New data could be used almost directly after data taking, because of a newly developed trigger scheme, called `Turbo`. During the first long shutdown and Run 2 of the LHC this trigger scheme and closely related `TurboCalib` was developed [114, 115]. For this purpose the full event reconstruction was needed. To do this, it is needed to align the sub-detectors and calibrate their output. It was made possible to perform this in an automatised way before the last trigger state, HLT2, by buffering the output after HLT1 and subsequently performing the full aliment and calibration before applying HLT2. Thus it was made possible to perform the same selections online in the HLT2 stage as offline. This allowed for saving only reconstructed objects of signal candidates and saving these directly on tape, instead of only saving raw detector information and performing the whole reconstruction again later on. As bandwidth and not event rate is the bottle neck for saving data, it was possible to for example select more charm decays, leading to the first observation of *CP* violation in charm [116]. Another consequence is that it was possible to directly save the reconstructed signal objects together with the raw detector information for calibration purposes, omitting later offline reconstruction and Stripping steps, making this calibration data available almost directly. This stream is called `TurboCalib` and is used for this analysis, which allows to directly select events with signal candidates and also run VELO reconstruction for probe tracks offline if necessary.

For the electron reconstruction efficiency measurement, a number of `TurboCalib` HLT2 trigger lines have been implemented to apply the method on $B^+ \to J/\psi K^+$, $B^+ \to J/\psi \pi^+$ and later for 2018 also lines for $B^0 \to J/\psi K^*$ and $B_s^0 \to J/\psi \phi$. To cover reconstruction efficiencies of other particle species, similar lines have been added for $J/\psi \to \mu^+\mu^-$ and $J/\psi \to p\overline{p}$.

At the base of the selection lie the tag particles, which are made up of a charged kaon and an electron. As they should come from the same vertex that is displaced from the PV, it is required that the tracks have a minimum $\chi^2/\text{ndf}_{\text{IP}}$ of 16. This is the difference in χ^2/ndf of including a track in the PV vertex fit or not. Both tracks also have to have a track χ^2/ndf of at most 5 to ensure a good track quality. To increase purity of the particle species, a particle-identification requirement has been set for the kaon at ProbNNk > 0.2 and for the electron at ProbNNe > 0.2. A minimum p_{T} of $500\,\text{MeV}/c$ for the kaon and $2.5\,\text{GeV}/c$ for the electron has been set, as backgrounds have generally lower momenta. For the electron, this has been set considerably higher, as this also approximately coincides with `L0` electron trigger requirements. For the electron, η has to be between 1.8 and 3.5, making the efficiency of the reconstruction of bremsstrahlung higher. This cut has been removed for 2018 data taking, to increase the overall efficiency and increase kinematic coverage.

On the combination of the tag electron and kaon some requirements have been set to reduce combinatorial backgrounds. These are combinations of tracks that do not originate from the same decay and consequently not the same vertex. A maximum on vertex χ^2,

Table 9.3: Selection requirements of the `Hlt2TrackEffElectronDetachedEEKTurboCalibLine` trigger line, used for the tag-and-probe method with 2017 data.

Candidate	Variable	Value	Unit
Event	$\text{nTracks}_{\text{HLT2}}$	< 120	
	`L0Electron`	`DEC`	
	`Hlt1TrackMVA`	`DEC`	
tag $[e^+]$	$\chi^2/\text{ndf}_{\text{track}}$	< 5	
	$\chi^2/\text{ndf}_{\text{IP}}$	> 16	
	η	$\in (1.8,\ 3.5)$	
	ProbNNe	> 0.2	
	p_{T}	> 2500	MeV/c
tag $[K^+]$	track χ^2/ndf	< 5	
	IP χ^2/ndf	> 16	
	ProbNNk	> 0.2	
	p_{T}	> 500	MeV/c
tag $[e^\pm K^+]$	χ^2_{vertex}	< 15	
	χ^2_{FD}	> 100	
	DIRA	> 0.95	
B^+ $[e^+e^-K^+]$	Mass	$\in (5000,\ 5700)$	MeV/c^2
	$\chi^2/\text{ndf}_{\text{vertex}}$	$< 20/3$	

χ^2_{vertex}, of 15 has been set. The vertex has to have a flight distance χ^2, χ^2_{FD}, of at least 100. This is calculated by subtracting the χ^2 of the combined vertex of the PV plus the tag combination minus the vertex χ^2 of the individual vertices of the PV and the tag combination, ensuring the displacement of the tag combination. In addition, a requirement on the pointing of the combination towards its assigned PV has been set by requiring the cosine of the direction angle, DIRA, to be at least 0.95. The direction angle is defined by the angular difference between the vector formed by the PV to the combination's vertex and the direction of the momentum of the combination in the lab frame.

The tag particles are combined with a VELO track and the resulting combination has to have a fairly loose requirement on its vertex's χ^2/ndf of at most 20/3. With the momentum of the probe calculated with eq. (9.2), an invariant mass of the full combination can be calculated and a range from $5000\,\text{MeV}/c^2$ to $5700\,\text{MeV}/c^2$ is required.

Next to these requirements at `HLT2`, requirements at `L0` and `HLT1` level have been set. For `L0`, it is required that `L0Electron` has fired in the event in general, called `DEC`, *i.e.* not necessarily requiring the electron to fire this. This is done with the purpose that this determination can be done offline, as it is crucial that the event is *not* fired by the probe electron. For `HLT2` the `Hlt1TrackMVA` has to be fired, which basically requires a track with a minimum impact parameter significance and a minimum p_{T} on the level of a few GeV/c, depending on the impact parameter significance and data taking.

The above-described selection criteria of the `HLT2` trigger line used for this analysis, `Hlt2TrackEffElectronDetachedEEKTurboCalibLine`, are summarised in table 9.3.

It should be noted that, although one cannot go back in time and change the trigger,

it is still possible to apply the method with a lower efficiency in previous data taking years, but still, as required for the method, not relying on the reconstruction of the probe.

9.4.2 Offline

As the selection on the trigger for the electron probe track is still quite loose, the purity of the data is quite low. Therefore, further selection on the probe is necessary. The advantage of moving this further selection offline is that one can reapply the selection, since selection on the probe is *a priori* more dangerous.

First of all, the η of the probe VELO track is required to be within the range of the kinematic binning. Second, the probe has to have an impact parameter, IP, of at least 0.2 mm. It was chosen not to use $\chi^2/\text{ndf}_{\text{IP}}$ to be independent on the estimation of the uncertainties of the track. Further, as a fiducial requirement, it is ensured that the probe is not a clone of the tag particles by requiring a minimum opening angle between the probe and the individual tag particles, in the lab frame, by having a maximum $\cos(\theta^{\text{lab}})$ of 0.9998.

By using the `DecayTreeFitter`, invariant masses of the J/ψ and B^+ candidate have been calculated by constraining the momentum of the B^+ to point from the PV. These have been used to set highly efficiency requirements on their ranges by requiring the J/ψ mass, $m^{\text{DTF;PV}}_{e^+e^-}$, the B^+ mass, $m^{\text{DTF;PV}}_{e^+e^-K^+}$, and the difference between the B^+ and J/ψ mass, $m^{\text{DTF;PV}}_{e^+e^-K^+} - m^{\text{DTF;PV}}_{e^+e^-}$, to be within the range of $(1400, 6000)$, $(3000, 7500)$ and $(1500, 3200)\,\text{MeV}/c^2$ respectively. Next to these, the invariant mass used for the likelihood fit to extract the efficiency, $m_{J/\psi K^\pm}$, has to be within the fit range of $(5050, 5650)\,\text{MeV}/c^2$. It is calculated with the `DecayTreeFitter` with a constraint on e^+e^- to have the invariant mass of the J/ψ. Apart from invariant-masses, one can use the momenta of the particles obtained with the `DecayTreeFitter` together with the decay vertex and PV to calculate the cosine of the direction angle, called DIRA^{DTF}. It is required that $\text{DIRA}^{\text{DTF}} > 0.999985$.

For the event not to be triggered by the probe electron, the tag electron has to fire `L0` and the tag electron or kaon `HLT1`, by requiring `TOS` for the particles for the used lines.

The offline selection is summarised and listed in table 9.4.

9.4.3 Matching

Aside from selecting the VELO probe track and the tag particles, to measure the efficiency for the probe to have a long track associated to it, it is needed to be able to select this long track out of all the long tracks that are reconstructed. Naively, matching of a VELO track to a long track is straightforward as one could equate the VELO track of the long track to the VELO track used as a probe. Nevertheless, due to subtleties in reconstruction, as described in section 8.1, these are not necessarily the same. This comes down to the fact that the overlapping part is at the input of the track building, not the output. The input for both is from the pattern recognition of the VELO, the VELO seeds. Due to the track fit, done separately on the VELO seed for the VELO track or on the entire long-track candidate, hits might be left out in one or the other. For the same reason, one might pass track quality requirements, while the other one does not.

Accordingly the requirement of at least 70% and not 100% of hits that are overlapping is used. The same value is used for the hit overlap fraction for truth-matching criteria in simulation. In addition, it is desirable that the overlapping hits on itself have the

Table 9.4: Definition of the offline selection, beside the trigger selection defined in table 9.3.

Candidate	Variable	Value	Unit
probe	η	$\in (1.9, 4.5)$	
	IP	> 0.2	mm
tag $[e^+]$	`L0Electron`	`TOS`	
tag $[e^\pm K^+]$	`Hlt1TrackMVA`	`TOS`	
probe/tag $[e^\pm K^+]$	$\cos(\theta^{\text{lab}})$	< 0.9998	
J/ψ $[e^+e^-]$	$m^{\text{DTF;PV}}_{e^+e^-}$	$\in (1400, 6000)$	MeV/c^2
	$\cos(\theta^{\text{lab}}_{e^+e^-})$	< 0.9998	
B^+ $[e^+e^-K^+]$	FD	> 4	mm
	DIRA^{DTF}	> 0.999985	
	$m^{\text{DTF;PV}}_{e^+e^-K^+}$	$\in (3000, 7500)$	MeV/c^2
	$m^{\text{DTF;PV}}_{e^+e^-K^+} - m^{\text{DTF;PV}}_{e^+e^-}$	$\in (1500, 3200)$	MeV/c^2
	$m_{J/\psi K^\pm}$	$\in (5050, 5650)$	MeV/c^2

minimum requirement to form a VELO track, *i.e.* at least three stereo VELO hits. Due to the nature of the decay that is used for this study, another requirement is possible by having the charge of the long track to be opposite to the charge of the tag electron.

With these criteria, the efficiency of reconstructing an electron as a long track given it is reconstructed as a VELO track is given by the ratio of the number of signal electron VELO tracks having at least one long track that matches out of all long tracks in the event over the total amount of signal electron VELO tracks.

Any systematic uncertainty arising from the matching criteria is taken into account and will be discussed in section 9.8.1.

9.5 Efficiency extraction: likelihood fits

To extract the reconstruction efficiency from a dataset, one has to determine the yield of signal probe-tracks that have a matching long-track, called a pass, and the yield of the ones that don't, called a fail. To distinguish signal from background, the invariant mass of the $e^+e^-K^+$ combination, $m_{J/\psi K^\pm}$, is used, where the e^+e^- combination is constrained to have an invariant mass of the J/ψ. An unbinned extended maximum-likelihood fit is used to directly extract the efficiency by maximising a combined likelihood of pass and fail categories [16, 65]. The signal yields per category, which are used for the Poissonian term in the extended likelihood, are given by,

$$\nu_{\text{pass}} = \nu_s \epsilon + \nu_{b\text{-pass}} \qquad \text{and} \qquad \nu_{\text{fail}} = \nu_s (1 - \epsilon) + \nu_{b\text{-fail}} \ , \tag{9.5}$$

where ϵ is the efficiency, ν_s the total signal yield and ν_{pass} and ν_{fail} are the signal yields in the pass and fail categories respectively. The parameters $\nu_{b\text{-pass}}$ and $\nu_{b\text{-fail}}$ are the number of background events in the pass and fail categories respectively. With this, the full

extended likelihood is given by,

$$\mathcal{L} = \frac{(\nu(1-\epsilon)+\nu_{b\text{-fail}})^n e^{-\nu(1-\epsilon)-\nu_{b\text{-fail}}}}{n!} \prod_i^n f_{\text{fail}}(m_{J/\psi K^\pm}|\vec{\lambda}) \tag{9.6}$$
$$\times \frac{(\nu\epsilon+\nu_{b\text{-pass}})^m e^{-\nu\epsilon-\nu_{b\text{-pass}}}}{m!} \prod_j^m f_{\text{pass}}(m_{J/\psi K^\pm}|\vec{\lambda}) \ ,$$

where f_{pass} and f_{fail} are the PDFs for the pass and fail category respectively and $\vec{\lambda}$ is the vector of parameters of the PDFs.

To determine f_{pass} and f_{fail}, the composition of backgrounds has to be known. Apart from the signal, the invariant-mass spectrum contains combinatorial backgrounds and partially-reconstructed backgrounds from $B \to J/\psi(\to e^+e^-)K^+\pi$, *i.e.* both B^0 and B^+, as the missing mass of a pion falls within the invariant-mass range. Another background comes from $B^+ \to J/\psi(\to e^+e^-)\pi^+$, but results in a negligible contribution due to a relative branching fraction of 3.8% with respect to the $B^+ \to J/\psi(\to e^+e^-)K^+$ decay in combination with a particle-identification requirement on the kaon which reduces it by about an order of magnitude. Also, it is no contamination of the signal, since it still contains $J/\psi \to e^+e^-$. Therefore, it is to first order independent of matching category, resulting in a negligible contribution to the efficiency determination. To conclude, the full PDFs, f_{pass} and f_{fail}, are described by a sum of the PDFs of the signal, $B^+ \to J/\psi(\to e^+e^-)K^+$, the combinatorial backgrounds and the partially-reconstructed backgrounds, $B \to J/\psi(\to e^+e^-)K^+\pi^-$.

The PDF for the signal is described by a double-sided Crystal Ball function, as defined in the Appendix by eq. (A.1). Essentially, it has a Gaussian core with power-law tails on both sides to account for a tail of higher invariant-mass uncertainties. The uncertainty on the invariant mass is positively correlated with the momentum of the decaying particle and subsequently the momentum of the probe. Subsequently, it is necessary to have separate values for the parameters of the PDF per p_{T} bin. The kinematic bins have a high enough statistical power to let the mean and width of the Gaussian core free in the fit. In contrast, this does not hold for the tail parameters and therefore they are fixed in the fit. They are determined from a likelihood fit to truth-matched simulation per p_{T} bin. The PDF is chosen to be the same for the fail as for the pass category, since the invariant mass is calculated independently from whether the electron should have a long track associated to it or not and any momentum effect is caught by binning the tail parameter determination per p_{T} bin. Any systematic arising from this choice is taken into account.

Due to the slight mass-uncertainty-momentum correlation in combination with the different invariant-mass requirements in the trigger and offline analysis, the edges of the invariant-mass spectrum are suppressed at high p_{T} of the probe. Consequently, the combinatorial backgrounds are not exponential any more. Therefore, the use of a second-order polynomial is used. To ensure that the polynomials are positive in their whole range, both in terms of invariant mass and parameter space, it was chosen to use Bernstein polynomials [117]. All the parameters of the polynomials are left free in the fit and are separate for pass and fail category, because a priori there was no reason found why the composition of the combinatorial backgrounds should be the same for the categories.

The partially-reconstructed backgrounds of $B \to J/\psi(\to e^+e^-)K^+\pi$ are described by an ARGUS distribution that is convoluted with a resolution function based on the PDF of the signal. This is because the invariant-mass resolution is the same, since the decays

also contain a $J/\psi \to e^+e^-$ decay from a B. The ARGUS distribution is used to describe the continuous distribution caused by the missing momentum [118]. The definition of the distribution is given in eq. (A.5). It was also used to describe partially-reconstructed backgrounds in other LHCb analyses [54]. The missing four-momentum of the pion ensures that the kinematic threshold is at $m_{B^+} - m_{\pi^0}$ or $m_{B^0} - m_{\pi^+}$ and as the difference of these two is small compared to the resolution, the value is fixed to the average, computed with values for the masses from the PDG [16]. The parameter determining the curve towards lower masses, *i.e.* the coefficient of the exponential factor, is left free in the fit. For the resolution function, the parameters are all the same as the signal PDF, except the mean, which is set to the mean of the signal PDF determined in a fit to simulation minus the B^+ mass from the PDG [16]. The full PDF is the same for fail and pass categories, chosen with the same reasoning as for the signal.

To build a combined PDF for the signal and background components, the fractions of the individual components of the PDF, both for f_{fail} and f_{pass}, are determined according to the individual yield divided by the sum of the individual yields. The individual yield of the signal is given by eq. (9.5). The yield of the other components are free parameters in the likelihood.

To increase the stability of the fit, priors have been added to ensure the likelihood maximum stays away from edges of the parameter space of the parameters of the combinatorial PDF and slope parameter of the ARGUS distribution used as PDF for the partially-reconstructed backgrounds. For this purpose, bifurcated Gaussian PDFs have been multiplied with the likelihood function of eq. (9.6). The values for the prior PDFs are listed in table 9.10. As this might introduce a bias, it is taken into consideration in the systematic uncertainty evaluation in 9.8.

The full likelihood function is maximised with the MINUIT package, using MIGRAD for the minimisation of $-\log(\mathcal{L})$ and HESSE for the determination of the uncertainties of the nuisance parameters and MINOS for the determination of the asymmetrical uncertainties of the efficiency [72, 73]. To assess if the likelihood fit is unbiased and estimates the right uncertainty, pseudo-experiments have been performed, which are discussed in section 9.8.2.

The likelihood fit to the full sample, both for simulation and for data, is shown in fig. 9.6 and fig. 9.7 respectively. Note that in simulation the partially-reconstructed background is not present and therefore not added to the full PDF. The dataset contains 92482 ± 513 $B^+ \to J/\psi(\to e^+e^-)K^+$ decays and the fitted PDFs show a visibly good agreement between simulation and data.

Figure 9.6: Distributions of invariant mass, $m_{J/\psi K^\pm}$, in simulation. The pass and fail categories are shown to the left and right respectively. The PDF of the fit is overlaid, including individual components, shown with dashed lines.

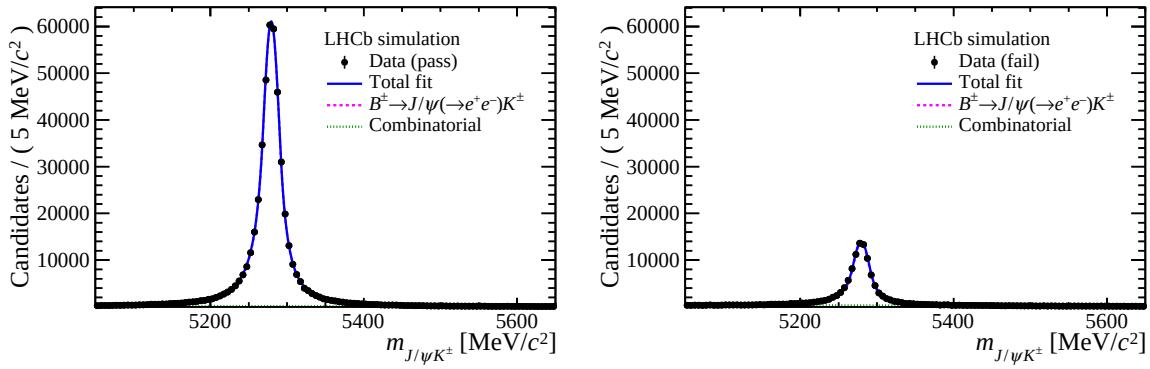
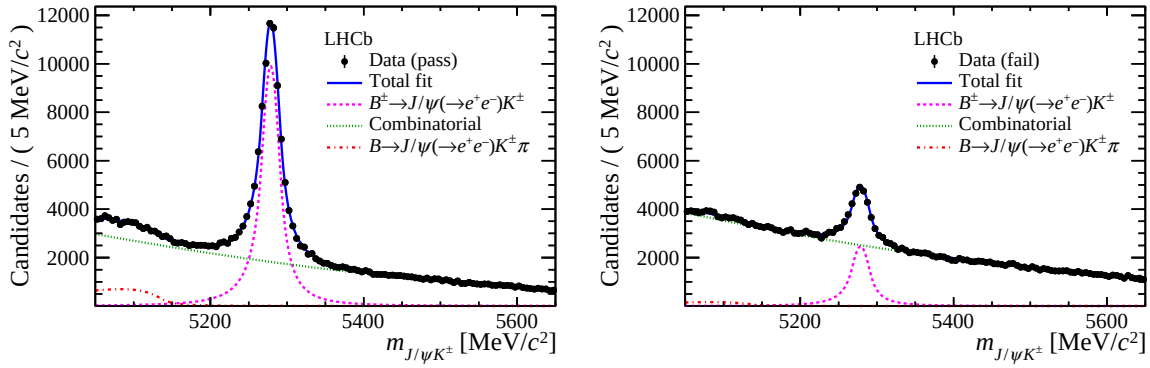


Figure 9.7: Distributions of invariant mass, $m_{J/\psi K^\pm}$, in data. The pass and fail categories are shown to the left and right respectively. The PDF of the fit is overlaid, including individual components, shown with dashed lines.



9.6 Ghosts

The likelihood fit only discriminates between signal and background VELO tracks. Hence, the match to the long track might still give a T-track that does not originate from the signal electron VELO track. To get a better understanding of the composition of such fake matches, called matching ghosts, simulation is used.

First, simulation is processed in the same way as data. Then, truth-matching information is retrieved from the reconstructed objects with links between hits and simulated particles. The truth matching is performed per track segment, *i.e.* separate per VELO track and T-track. By requiring the VELO track to come from the signal and looking at truth-matching information of the T-track, the origins of these matching ghosts are determined and listed in table 9.5.

The effective ghost-rate is about 2% and interestingly, the majority of the fake matches are signal related. Most of them come from a conversion of a bremsstrahlung photon coming from the track itself. The rest comes from a conversion of a final-state radiation photon and pions, kaons, protons, muons and electrons that are not related to the signal. Since these ghosts contain the right VELO but not the right T-track, the resulting long

Table 9.5: Origin of the T-track of the matched long-track in simulation. Fractions are relative to probe electrons with a matching long-track. Uncertainty on the fractions is about 0.01%. In case op bremsstrahlung and final state radiation, the origin of the T-track is an electron from the conversion of the respective photon. Sample is simulation of the $B^+ \to J/\psi(\to e^+e^-)K^+$ decay passing the `Hlt2TrackEffElectronDetachedEEKTurboCalibLine` trigger line.

Type of origin	Fraction	Fraction / Ghosts
signal	97.86%	
best match	95.99%	
ghosts	2.14%	
related to signal		
bremsstrahlung		
probe	1.55%	72.43%
tag	0.00%	0.00%
final-state radiation	0.07%	3.27%
other origins		
pions, kaon, protons, muons	0.33%	15.42%
other electrons	0.20%	9.35%
T-track ghosts	0.00%	0.00%
other origins	0.00%	0.00%

tracks do not have the correct momentum. Nonetheless, this can also happen with a non-ghost long-track due to bremsstrahlung emission. Due to this indistinguishability, these ghosts are seen as part of the signal. This is acceptable as long as one corrects for this in their efficiency determination and these ghosts are included in simulation. This is one of the reasons why the efficiencies have to be determined per specific selection of electrons. For example, a particle identification requirement might remove a lot of these ghosts by essentially requiring $E_{\text{cluster}}/p_{\text{track}}$ to be close to one. A fake T-track will also have the wrong electromagnetic calorimeter cluster associated to it, if it has one in the first place. This results in $E_{\text{cluster}}/p_{\text{track}}$ to be less likely closer to one.

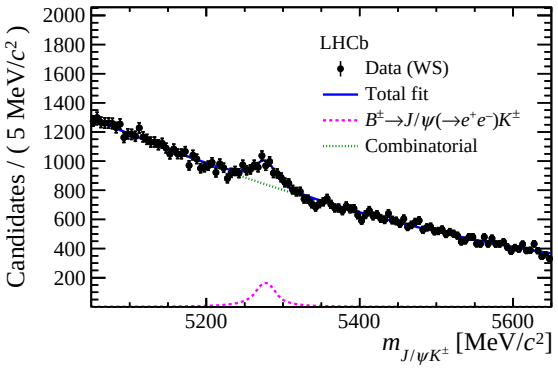
There are two ways to investigate these ghosts in data. First of all, one can look at a sample to matches that one knows for sure is not signal, *i.e.* not a true track. The easiest way is to match to a track that has a wrong charge. To compare wrong versus right-sign matching-ghosts, the simulation study of table 9.5 is performed for the wrong-sign matches and its results are listed in table 9.6. The rate of wrong-sign matches relative to right-sign is $(1.52 \pm 0.01)\%$, which is of the same order as the right-sign ghost rate of $(2.14 \pm 0.01)\%$. In data, one can extract the wrong-charge matches directly with the nominal fit method on the wrong-sign data sample as the signal VELO track has only wrong matches. This fit is shown in fig. 9.8 and, relative to the nominal data set, has a rate of $(1.88 \pm 0.27)\%$, which is compatible with the one found in simulation. Assuming the ratio of wrong over right-sign matches in simulation is the same as in data, the right-sign ghost-rate in data is $(2.64 \pm 0.38)\%$.

Second, one can look at the invariant mass determined with the momenta from the track fit to the matched long-tracks, $m^{\text{LL}}_{e^+e^-}$. Since the momenta of the matching ghosts

Table 9.6: Origin of the T-track of the matched long-track with wrong charge in simulation. Fractions are relative to probe electrons with a matching long-track with wrong charge. Rate of wrong-sign with respect to right-sign matches is $(1.52 \pm 0.01)\%$. Uncertainty on the fractions is about 0.1%. Same description and simulation sample as for table 9.5.

Type of origin	Fraction	Fraction / Ghosts
signal	0.06%	
best match	0.06%	
ghosts	99.94%	
related to signal		
bremsstrahlung		
probe	68.74%	68.78%
tag	0.01%	0.01%
final-state radiation	4.16%	4.16%
other origins		
pions, kaon, protons, muons	19.21%	19.22%
other electrons	7.99%	7.99%
T-track ghosts	0.09%	0.09%
other origins	0.00%	0.00%

Figure 9.8: Distributions of invariant mass in data for the wrong-sign sample. The amount of signal with respect to the signal yield of matched probes in the nominal dataset is $(1.88 \pm 0.27)\%$.



have a more random momentum than signal electrons, the invariant mass has a more flat distribution than signal electrons of which the distribution peaks at the J/ψ mass. Since $m^{\mathrm{LL}}_{e^+e^-}$ is orthogonal to the nominal fit because of the J/ψ-mass constraint, the `sPlot` technique can be applied [63] to obtain the distribution of $m^{\mathrm{LL}}_{e^+e^-}$ for signal probe VELO tracks with a long-track match. Therefore the weighted $m^{\mathrm{LL}}_{e^+e^-}$ distribution only contains long tracks with signal T-tracks and matching ghosts, which are the two components of which the yield has to be determined. The PDF of the matching ghosts is described by kernel density estimation [71]. It uses a sum of Gaussian PDFs where the mean of each Gaussian is given by a data point in a reference dataset, in this case truth-matched simulation. The reference dataset with the PDF overlaid is shown in fig. 9.9. For the

Figure 9.9: Distributions of invariant mass of the e^+e^- combination calculated with long-track momenta in simulation for signal (left) and ghosts (right), with fit overlaid.

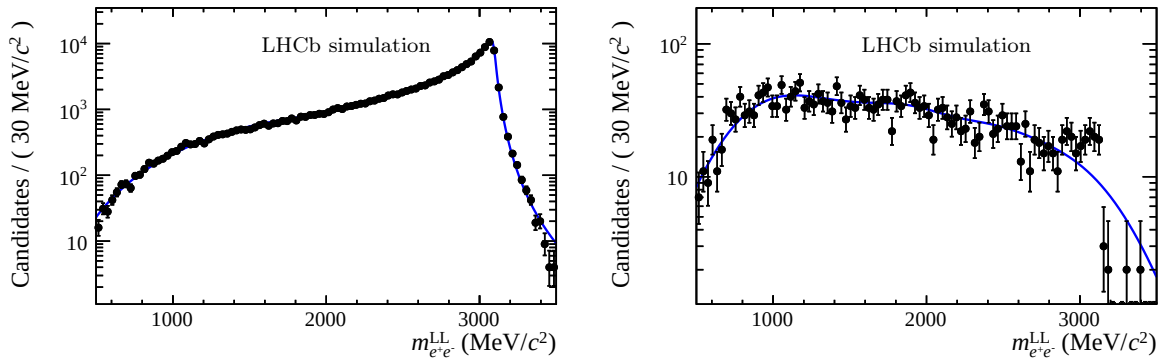
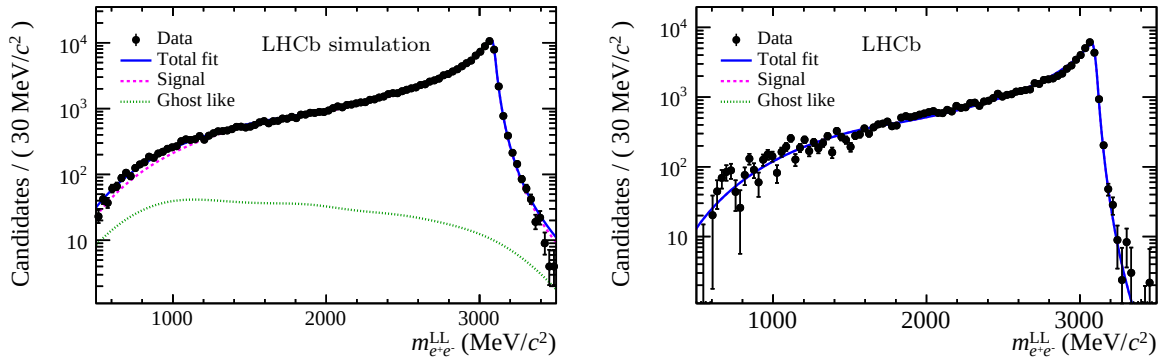


Figure 9.10: Distributions of invariant mass of the e^+e^- combination calculated with long-track momenta in simulation (left) and data (right), with fit overlaid. Note that the ghost contribution converged to zero in the data fit.



PDF of the signal, a modified version of the double-sided Crystal Ball, called thresholded double-sided Crystal Ball, is used and defined in eq. (A.6). The purpose is to add at the lower tail of the double-sided Crystal Ball another tail that is used to describe a kinematic threshold, described by a function that is linear in a double-log plot. The thesholded double-sided Crystal Ball PDF, fitted to the signal sample, is shown in fig. 9.9. A fit to the full long-matched simulation sample, *i.e.* with no truth-matching on the T-track, is shown in fig. 9.10, together with the fit to data. While the fit to simulation performs quite well, which should be, as the individual components are determined on this sample, the data fit, with less candidates, is less discriminant. It converged to a zero ghost-rate and therefore does not give a sensible value and uncertainty. To show there is still some discriminatory power, the PDF with fixed ghost-rates of 2% and 8% are shown in fig. 9.11. The one with 2% shows a similar agreement, while the one with 8% clearly does not. As there is no clear way to know if the PDF for the ghost contribution is similar in simulation and data, there is also no way of determining a systematic uncertainty on this. To conclude, the use of $m^{\rm LL}_{e^+e^-}$ does not have enough discriminatory power to obtain the ghost rate.

To summarise, the ghosts related to matching VELO tracks to long tracks, of which the majority come from signal-related bremsstrahlung conversions, are indistinguishable from real electrons loosing a lot of energy due to bremsstrahlung. In addition, this makes

Figure 9.11: Distributions of invariant mass of the e^+e^- combination calculated with long-track momenta in data with PDF overlaid with fixed ghost rates of 2% (left) and 8% (right).

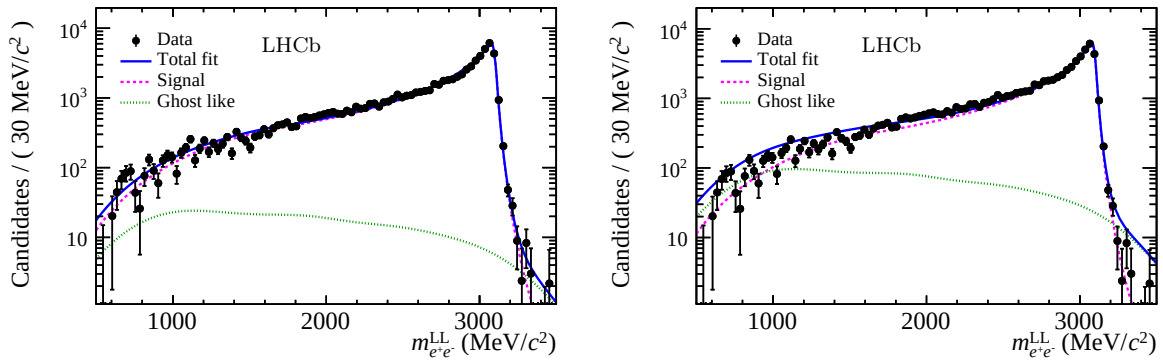


Table 9.7: Summary of ghost rates for wrong and right-sign matches in simulation and data. All rates are with respect to number of right-sign matches. Estimate of right-sign data ghost-rate is based on the assumption that the ratio between wrong and right-sign in simulation is the same as in data.

Type and dataset	Rate	Estimate
Wrong-sign		
data	$(1.88 \pm 0.27)\%$	-
simulation	$(1.52 \pm 0.01)\%$	-
Right-sign		
data	?	$(2.64 \pm 0.38)\%$
simulation	$(2.14 \pm 0.01)\%$	-

it hard to extract the ghost rate from data with an invariant-mass fit. Despite that, the nominal fit method can be used to extract wrong-sign matches and a comparison can be made with simulation and one can extrapolate to right-sign matches, estimating a rate of about 2% to 3%. The different ghost rates are summarised in table 9.7.

Due to the before-mentioned indistinguishability, the efficiencies are measured including these ghosts. It is advised to analysts to include ghosts in their analysis when using the efficiency ratios to correct their simulation. This essentially means that one should use simulation which includes these ghosts by not requiring truth-matching on the T-track segment. Hence, one should perform truth matching on the VELO segment at most.

9.7 Efficiencies: data and simulation

The electron reconstruction efficiencies extracted with the likelihood fit for the kinematic bins are shown in figs. 9.12 and 9.13, including the ratio. The kinematic binning is discussed in section 9.2 and defined in tables 9.1 and 9.2. All uncertainties are statistical. The uncertainty on the ratio is calculated assuming that the uncertainties between data and simulation are uncorrelated.

While there is overall good agreement between data and simulation, there are some regions with tensions. In the low η bin, at low p_{T}, the data seems less efficient than simulation. The figures of the fits in this region with their PDFs and the data per pass and fail categories are shown in appendix C.1. All fits have converged and show a good goodness-of-fit, shown with the pull of data from the fitted PDF.

As this suggests that the discrepancy lies at low momentum, the likelihood fits have been performed with the same kinematic binning, but with a global requirement of $p > 3\,\mathrm{GeV}/c$. The results of these fits for $1.9 < \eta < 2.9$ in the non RF-Foil region are shown in fig. 9.14. This shows that the disagreement between simulation and data becomes less at low p_{T}. Any further properties and the origin of this discrepancy are unknown. More data might reveal this, since the RF-Foil region seems to have less of discrepancy, although it has low statistical power. This might hint towards a difference originating from different amounts of material in data versus simulation downstream of the VELO. Nevertheless, for many analyses these low-momenta electrons are already removed by fiducial requirements and the efficiency ratios obtain with this method can be used to correct for this if this phase-space region is not already removed.

To be able to use the results of these fits, the systematic uncertainties have to be determined first, which are discussed in the following section, section 9.8. Then, the way these efficiency ratios can be used for corrections and systematic uncertainty determinations in analyses will be briefly discussed in section 9.9, including an example.

Figure 9.12: Efficiencies for the non-RF-Foil region. Ratio is data over simulation efficiency. All uncertainties are statistical, where for the ratio, the uncertainties on data and simulation are assumed to be uncorrelated. It shows an overall good agreement, only at low η and p_{T} the data efficiencies are lower.

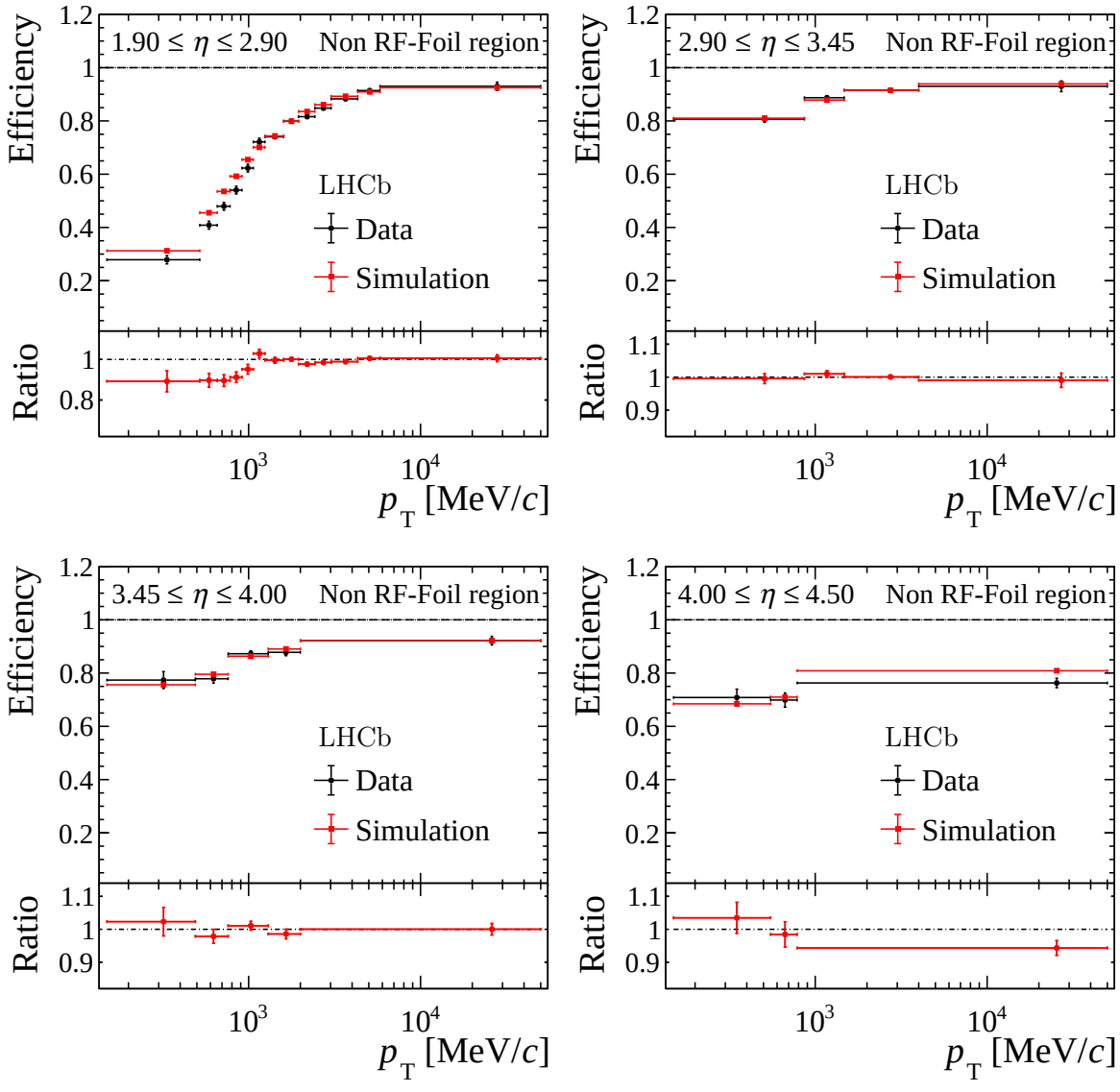


Figure 9.13: Efficiencies for the RF-Foil region. Ratio is data over simulation efficiency. All uncertainties are statistical, where for the ratio, the uncertainties on data and simulation are assumed to be uncorrelated. It shows an overall good agreement.

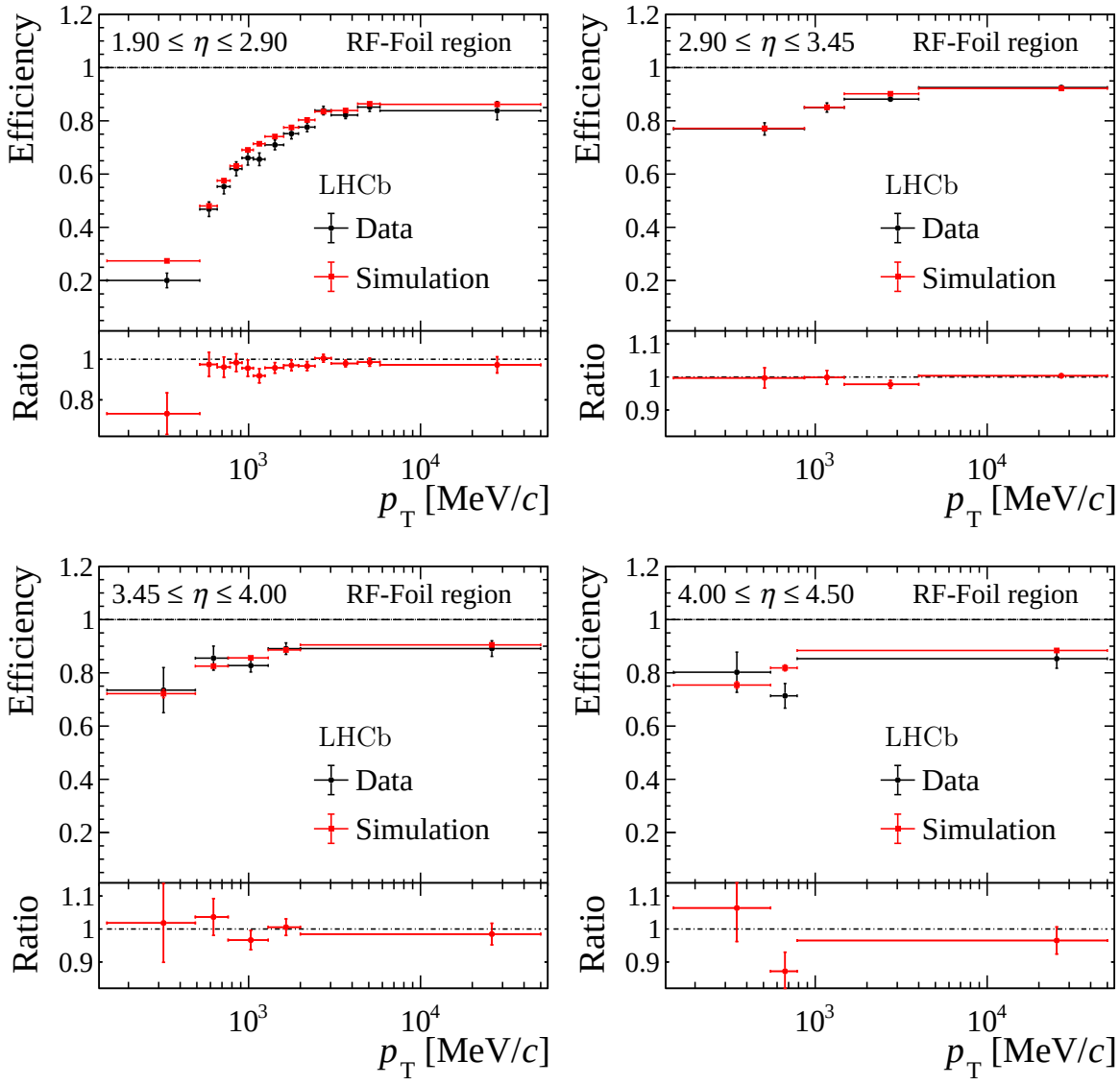
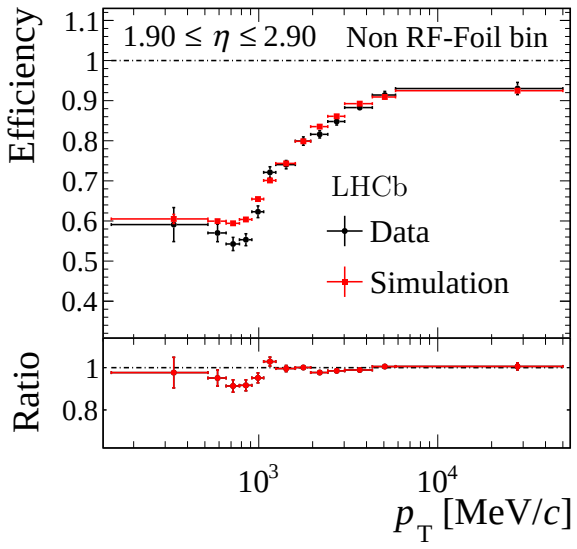


Figure 9.14: Efficiencies for the low η bin in the non-RF-Foil region with an minimal requirement of 3 GeV/c on the momentum of the probe. Differences at low p_{T} become less pronounced between data and simulation. Ratio is data over simulation efficiency. All uncertainties are statistical, where for the ratio, the uncertainties on data and simulation are assumed to be uncorrelated.



9.8 Systematic uncertainties

Since the main goal is to establish if there are any differences between simulation and data in electron reconstruction efficiency, the main concern is to have as low as possible systematic uncertainties on the ratio of data over simulation and not necessarily the absolute efficiencies themselves. Regarding the absolute efficiencies, biases due to the tag selection are of the order of a few percent, as shown and discussed in appendix C.2. This gives another reason to use the ratio as main result. Subsequently, the following systematic-uncertainty determinations are for the ratio.

Systematic uncertainties can arise in the following part of the analysis: the selection, including the matching criterium, fit model and momentum resolution. These are discussed in sections 9.8.1 to 9.8.3 respectively. The statistical and systematic uncertainties are combined and summarised in section 9.8.4.

In addition, stability cross-checks have been performed on data-taking conditions and detector occupancy. These are discussed in section 9.8.5.

9.8.1 Selection

Selection criteria can be split into two categories: those indirectly affecting the kinematic distributions of the probe, in this section called indirect criteria, and those based on variables using information from the probe directly. Therefore, they can affect the efficiency directly and are hereafter referred to as direct criteria.

In addition to unfolding the efficiencies of electrons from the kinematics of their decay, a part of the purpose of using kinematic binning in p_{T}, η and ϕ is to alleviate possible in-bin variations in kinematics between data and simulation. Indirect criteria can have an effect on this. As explained in section 9.2, the choice of binning is done in such a way that the bin size is small enough to account for large fluctuations in efficiency, while still having enough data to extract the efficiency. This ensures that statistical uncertainties dominate with respect to fluctuations of the efficiency and kinematic distributions. In addition, the differences between η and ϕ in simulation versus data are small, as can be seen in fig. 9.15. Transverse momentum, the remaining kinematic variable, will be separately discussed in section 9.8.3. All in all, no separate handling of indirect criteria is needed.

Direct criteria come into the selection at two levels: in the trigger and at the offline

Figure 9.15: Distributions of η and ϕ of the probe in data and simulation. The `sPlot` technique is applied to obtain the data distribution [63].

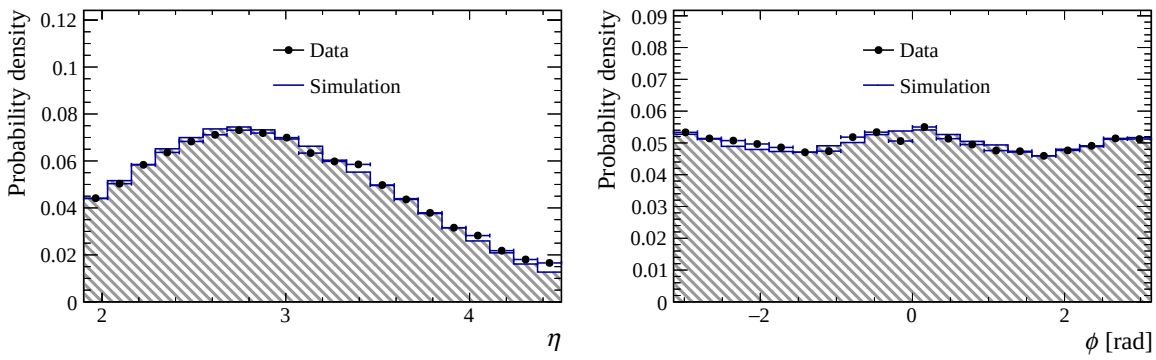


Table 9.8: Efficiencies of direct criteria conditional on the trigger selection of variables that involve information from the probe. Sorted with increasing efficiency. Full selection is given in tables 9.3 and 9.4. Determined with simulation. Uncertainties are negligible.

Candidate	Selection variable	Efficiency
B^+ $[e^+e^-K^+]$	$\mathrm{DIRA}^{\mathrm{DTF}}$	75.7%
	FD	80.7%
probe	IP	83.6%
combined mass requirements		96.7%
B^+ $[e^+e^-K^+]$	$m^{\mathrm{DTF;PV}}_{e^+e^-K^+}$	97.6%
J/ψ $[e^+e^-]$	$m^{\mathrm{DTF;PV}}_{e^+e^-}$	97.6%
B^+ $[e^+e^-K^+]$	$m^{\mathrm{DTF;PV}}_{e^+e^-K^+} - m^{\mathrm{DTF;PV}}_{e^+e^-}$	98.3%
B^+ $[e^+e^-K^+]$	$m_{J/\psi K^\pm}$	99.4%
combined opening angle requirements		99.0%

stage. In case of the trigger, two criteria are direct: the χ^2_{vertex} of the B and the mass range given by the simplified J/ψ-mass constraint. For the offline stage, the direct criteria are listed in table 9.8, together with their efficiencies conditional on the trigger selection.

These efficiencies give also, in part, their importance. The criteria with the lowest efficiencies are $\mathrm{DIRA}^{\mathrm{DTF}}$, FD and IP and therefore have to most potential impact on the reconstruction efficiency. For the mass requirement, both from the trigger and offline, it has been checked that any efficiency dependence is much smaller than for aforementioned variables. In addition, the efficiencies of the mass requirements are high and therefore no systematic uncertainty is assigned for the mass requirements.

The distributions of χ^2_{vertex}, $\mathrm{DIRA}^{\mathrm{DTF}}$, FD and IP are shown for data and simulation in fig. 9.16. It shows that χ^2_{vertex} has the largest difference between data and simulation. The focus is therefore on χ^2_{vertex} and will be discussed first. Second, the dependence on the remaining variables $\mathrm{DIRA}^{\mathrm{DTF}}$, FD and IP is discussed.

From simulation it has been determined that the efficiency dependence of χ^2_{vertex} is consistent among kinematic bins. Subsequently, to assess the dependence of the efficiency due to requirements on χ^2_{vertex}, likelihood fits over the full kinematic range have been performed for different values of the requirement on χ^2_{vertex} on top of the nominal selection, both in data and simulation. The result of these fits are shown in fig. 9.17. The most striking result is that in the ratio the strong dependence as seen for the absolute efficiency on the requirement value is gone. Thereby, the power of using the ratio (in the first place) is evident.

Despite that, since the nominal value of the χ^2_{vertex} requirement cannot be lifted in data and for simulation this is difficult, an estimation is made of the bias on the absolute efficiency by extrapolation. As the distribution of χ^2_{vertex} behaves exponentially for high values, the amount of added data to the fits for looser values of χ^2_{vertex} gets exponentially smaller. Subsequently, the systematic effect goes down exponentially. Therefore, to extract the bias of $\chi^2_{\mathrm{vertex}} < 20$ in the trigger, it is estimated by extrapolating a fitted efficiency

Figure 9.16: Distributions of χ^2_{vertex}, DIRA^{DTF}, FD and IP in data and simulation. Distribution in data is obtained with the `sPlot` technique [63].

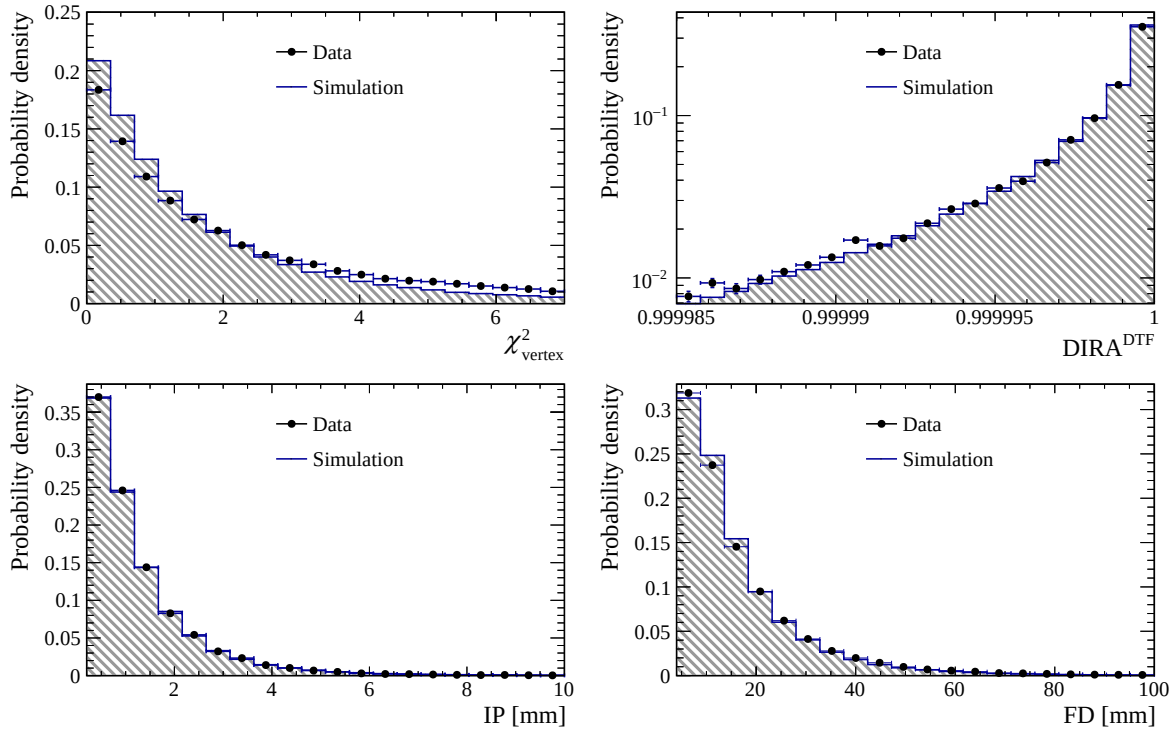
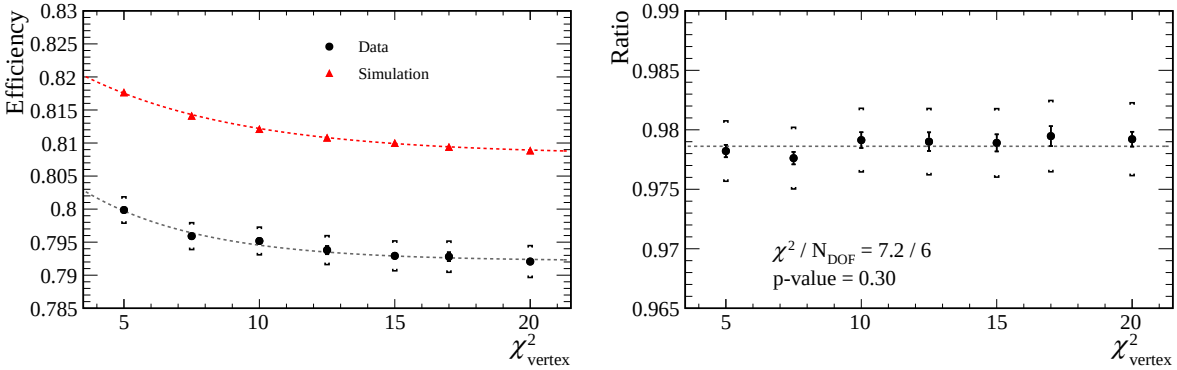


Figure 9.17: Absolute (left) and ratio (right) of reconstruction efficiencies measured integrated over kinematics for data and simulation with different upper bounds for χ^2_{vertex}, as denoted on the x-axis. The absolute efficiencies show a significant dependency, whereas the ratio does not with a precision of about 0.01%. Uncertainties denoted by the brackets are the full fit uncertainty, whereas the smaller uncertainties denoted by the vertical lines are the uncorrelated ones, determined by the squared difference between neighbours.



dependency, $\epsilon(\chi^2_{\text{vertex}})$, by,

$$\epsilon(\chi^2_{\text{vertex}}) = \epsilon_0 + \delta\epsilon = \epsilon_0 + c_0 e^{-c_1 \cdot \chi^2_{\text{vertex}}} \tag{9.7}$$

where ϵ_0 is the estimated efficiency without χ^2_{vertex} requirement, $\delta\epsilon$ the efficiency difference at a certain value of χ^2_{vertex} and c_0 and c_1 are empirical constants. Determined by $\delta\epsilon$, the uncertainty on the efficiency is $(0.0732 \pm 0.0050)\%$ in simulation and $(0.024 \pm 0.025)\%$

Table 9.9: Changed requirements used for systematic uncertainty determination, compared to nominal selection. Full nominal selection is given in tables 9.3 and 9.4.

Candidate	Variable	Nominal value	Loosened value	Unit
probe	IP	> 0.2	> 0.15	mm
B^+ $[e^+e^-K^+]$	DIRA^{DTF}	> 0.999985	> 0.999975	
	FD	> 4	> 2	mm

in data. To be conservative, this small systematic uncertainty on the absolute efficiency is used as systematic uncertainty on the ratio. As the behaviour is consistent among kinematic bins, the fitted systematic uncertainty from simulation, as it is the largest, is used for the efficiency in each bin both for data and simulation.

For the remaining direct criteria, DIRA^{DTF}, FD and IP, the behaviour of the efficiency for different kinematic bins is not as consistent as for χ^2_{vertex}. Accordingly, the systematic uncertainties are determined per kinematic bin. If one would apply the same approach as for χ^2_{vertex}, *i.e.* including fits in data, but for each kinematic bin, the statistical uncertainties would be too high and the result would be too conservative. Therefore, it is chosen to determine the bias from the absolute efficiency in simulation. The uncertainty is determined from the difference between the nominal fits and fits with a loosened selection on these variables, given in table 9.9. The resulting uncertainties are low compared to the statistical uncertainty.

For the matching criterium, the systematic uncertainty is determined by performing and comparing the likelihood fits in kinematic bins with the nominal criterium, with 70% overlap of hits for the VELO track and the long track, to the more strict criterium of 100% overlap. The resulting uncertainty is consistent among bins and on average below 10% of the statistical uncertainty and about 0.1% on the absolute efficiency.

The systematic uncertainties determined for χ^2_{vertex}, DIRA^{DTF}, FD, IP and the matching criterium are combined and will be reported in section 9.8.4.

9.8.2 Likelihood fit

Systematic uncertainties due to the likelihood fit can arise in the following parts: the choice of PDF parametrisation, including the fixing of parameters from simulation, using priors for parameters and the likelihood maximisation.

The choice of the PDF parametrisation, or rather assumption, is that the distributions for pass and fail category are equal. The efficiency is especially sensitive to possible differences between these categories. To study this, the $m_{J/\psi K^\pm}$ distributions of the signal mode are fitted separately for the pass and fail categories with a double-sided Crystal Ball. These fits are performed in bins of probe electron p_{T}, since the distributions depends most on this variable. The differences for the mean and width of the core Gaussian of the double-sided Crystal Ball are shown in fig. 9.18. The results for the tail parameters are shown in appendix C.3. For all parameters, no differences are found between categories. As a conservative systematic uncertainty estimate, the differences in fit results in simulation between using the same PDF for pass and fail and using separate ones is used.

Aside from directly fixing variables from simulation, some priors have been added to

Figure 9.18: Difference between the pass and fail category of the mean, $\Delta\mu$, and width, $\Delta\sigma$, of the core Gaussian of the double-sided Crystal Ball used for the signal PDF, determined in simulation depending on p_{T} of the probe. No significant difference between the two categories is observed.

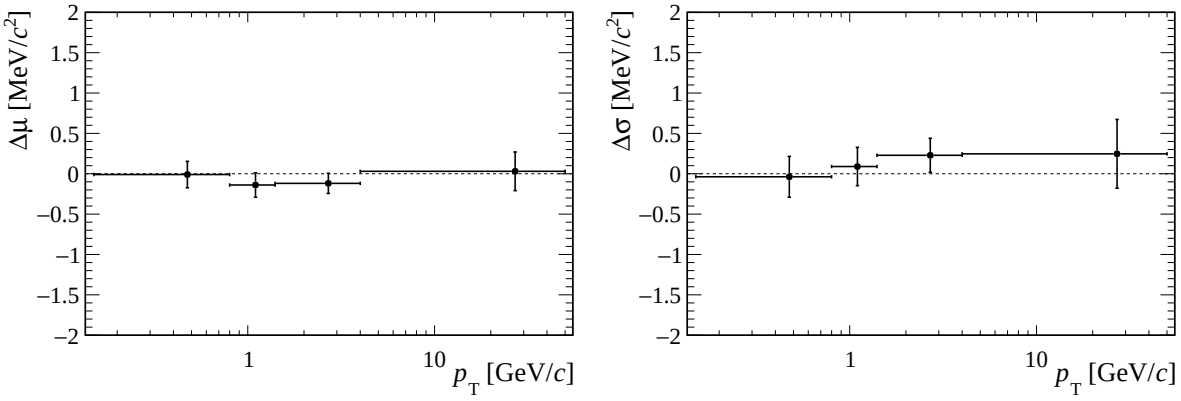


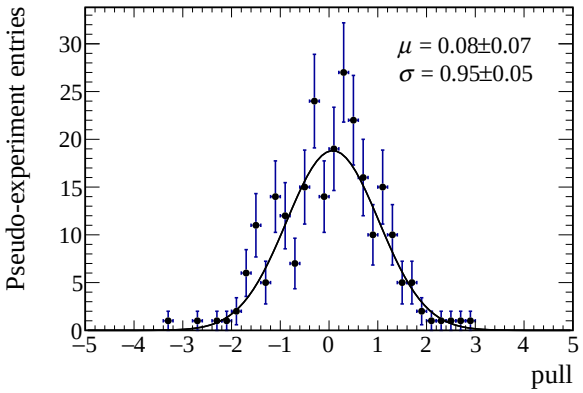
Table 9.10: Constraints on the likelihood to enhance the stability of the fit. The PDF to describe the prior PDFs is a bifurcated Gaussian. Listed are parameters that have these constraints and the parameters of the bifurcated Gaussian of the corresponding constraint.

Background	Category	PDF	Parameter	μ	σ (left)	σ (right)
Combinatorial	fail	Bernstein	c_1	0.15	0.25	0.70
		Bernstein	c_2	0.15	0.25	0.40
Combinatorial	pass	Bernstein	c_1	0.05	0.10	0.90
		Bernstein	c_2	0.08	0.16	0.58
$B \to J/\psi(\to e^+e^-)K^+\pi$		ARGUS	c	0	500	0.5

constrain some parameters in the likelihood [67]. The purpose is to enhance the stability of the likelihood maximisation and the determination of the uncertainties. This has been done for some of the parameters of the PDFs of the combinatorial and partially reconstructed backgrounds. Especially in bins with low statistics, this enhances stability. The constraints that have been applied are listed in table 9.10. The bias of introducing these constraints are determined to be maximally 4% with respect to the statistical uncertainty out of all the kinematic bins. This case is equal to a 0.11% absolute uncertainty. The resulting systematic uncertainty is therefore negligible.

For the likelihood maximisation, it has been checked that it returns both an unbiased central value and the right uncertainty. By generating a number of pseudo-experiments with the PDF obtained from fitting the full data sample, the response of the fit can be assessed. Knowing the input value of the efficiency, the fitted efficiency of each pseudo-experiment can be compared. The pull, *i.e.* the difference between input and fit output over the estimated uncertainty, can be obtained for this set of 250 pseudo-experiments, shown in fig. 9.19. For an unbiased central value, the mean of the pull distribution should be consistent with zero and the width should be one. As uncertainties are assumed to be Gaussian, the distribution of the pull should be Gaussian. Subsequently, the distribution of the pull is fitted with a Gaussian. The p-value of the χ^2 is 0.38 and the obtained mean

Figure 9.19: Distribution of the pull of the efficiency from pseudo-experiments obtained from the PDF resulting from the data fit to the full kinematic range. Distribution is consistent with a Gaussian, having a p-value of the χ^2 of the fit with a Gaussian of 0.38. No bias is observed and the uncertainty is estimated well, as the mean and width are consistent with zero and one, respectively.



and width of the distribution are consistent with zero and one, respectively. Hence, the likelihood maximisation of the fit model is performing well and no systematic uncertainty is assigned.

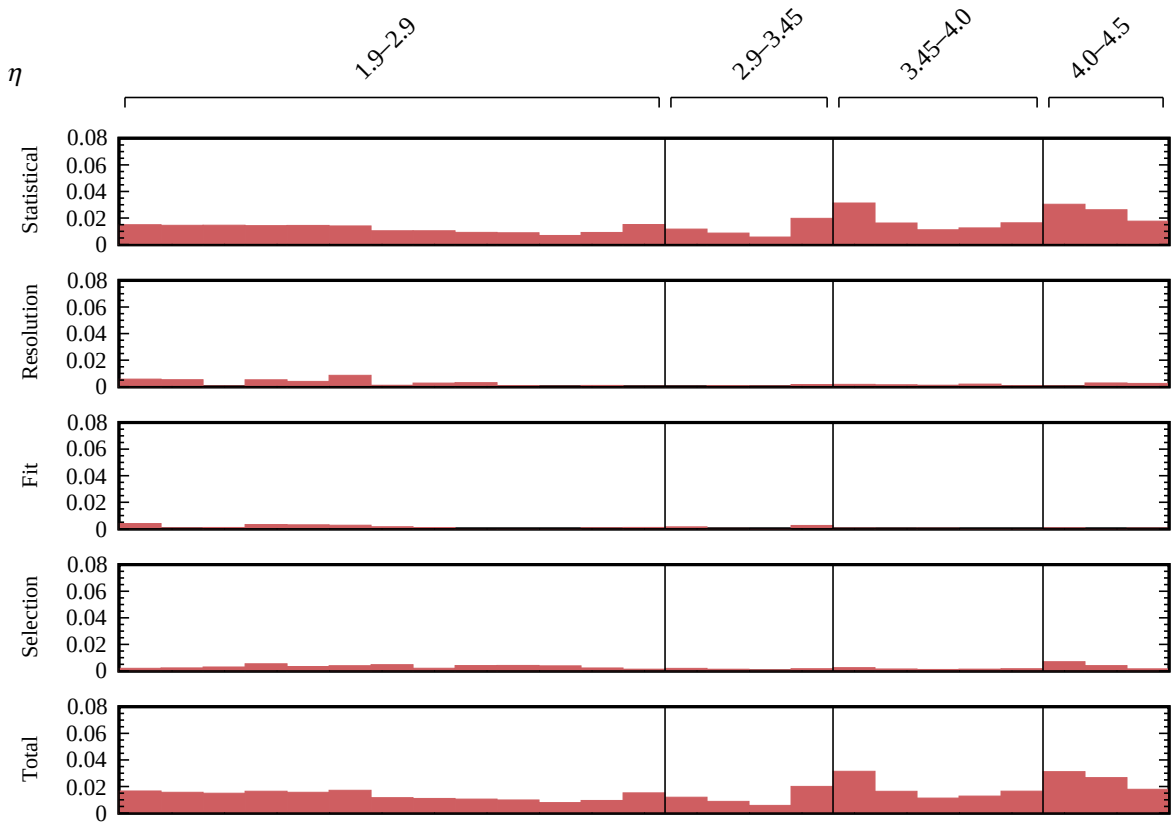
9.8.3 Momentum resolution

Aside from systematic uncertainties on the efficiency in general, the efficiency is determined for specific kinematic bins. If there is bin migration due to resolution effects, a systematic uncertainty has to be assigned. While η and ϕ are known to negligible precision, the momentum is not, as shown in section 9.3.

The effect on the absolute efficiency by bin migration is first of all proportional to the relative number of entries that migrate. Second, it is proportional to the efficiency difference between the bins where the event originated and ended up. Since if there is no difference, there is also no bias. In simulation, the difference between the efficiencies determined by binning with the true momentum (*i.e.* without counting final-state radiation and resolution effects) and the reconstructed momentum is determined. It shows absolute biases up to 1.6%, where the biases are largest in the kinematic ranges where the efficiency differs most, as expected.

For the efficiency ratio, the systematic uncertainty is proportional to the efficiency-ratio difference, similar to the absolute efficiency. Its calculation is performed by using the efficiency ratios as obtained from the likelihood fit and subsequently smearing them with the momentum resolution, shown in fig. 9.5, across the neighbouring kinematic bins. The smeared efficiency-ratios per bin are compared to the nominal efficiency-ratios and the difference is assigned as systematic uncertainty. The uncertainty is 0.15% on average and goes up to 2.5% due to the large statistical uncertainties and efficiency-ratio differences in the low-η, low-p_{T} range.

Figure 9.20: Uncertainties, both statistical and systematic, on the efficiencies ratios for the non-RF-Foil region, presented in fig. 9.12. The systematic uncertainties from the momentum resolution, likelihood fit and selection are reported separately. The bins for η are denoted at the top. Per η bin, the p_{T} bins are indexed with increasing p_{T}. It shows that the statistical uncertainties dominate the total uncertainties.

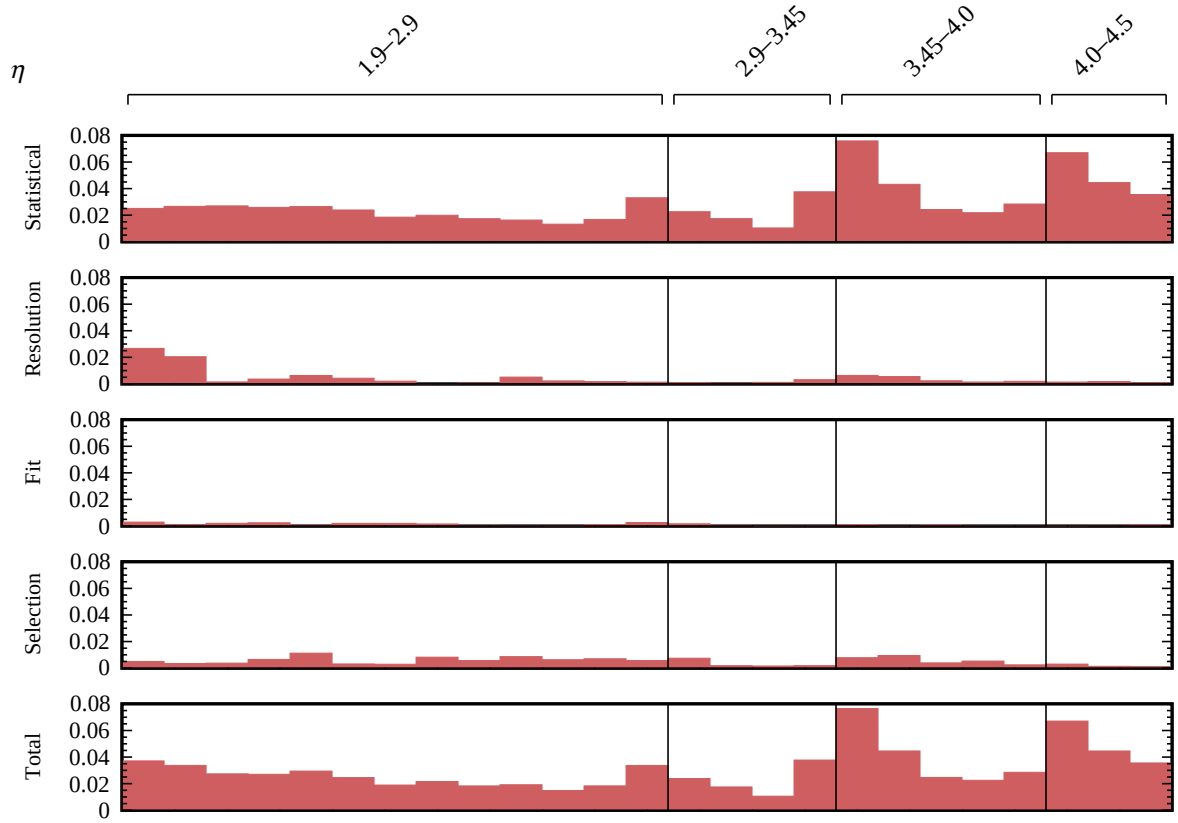


9.8.4 Total uncertainties

The complete set of uncertainties on the efficiency ratios, both statistical and systematic, are shown in figs. 9.20 and 9.21 for the non-RF-Foil and RF-Foil region respectively. The individual systematic uncertainties, as determined in sections 9.8.1 to 9.8.3 are shown as well.

In general, the systematic uncertainties are considerably lower than the statistical uncertainties and are on average 0.6%. Only in the low-η, low-p_{T} range the difference between data and simulation are considerable and the statistical uncertainties are large, especially in the RF-Foil region, leading to a sizeable systematic uncertainty due to bin migration due to momentum resolution. Note that this systematic uncertainty will go down with an increasing amount of data, as the systematic-uncertainty determination will be more precise.

Figure 9.21: Uncertainties, both statistical and systematic, on the efficiencies ratios for the RF-Foil region, presented in fig. 9.13. The systematic uncertainties from the momentum resolution, likelihood fit and selection are reported separately. The bins for η are denoted at the top. Per η bin, the p_{T} bins are indexed with increasing p_{T}. The uncertainties are dominated by the statistical one, aside from the low p_{T}, low η region. These are caused by the large statistical uncertainty in combination with the large efficiency changes in this region.



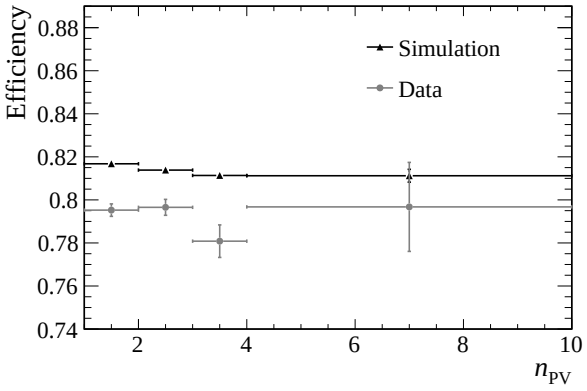
9.8.5 Stability cross-checks

Apart from measuring the efficiencies ratios themselves, it is relevant to see what dependencies the efficiency has, aside from purely kinematics. For example, detector conditions like occupancy and magnet polarity are of interest.

In the case of occupancy, for the muon tracking efficiencies, there is a dependency seen in the occupancy, specifically the multiplicity of hits in the SPD [119]. For electrons, the dependence on detector occupancy is checked as well. To be independent of decay kinematics, it is chosen to use the pile up, *i.e.* the number of PVs, as occupancy measure. It is determined for the full kinematic range for data and simulation and shown in fig. 9.22. It shows a slight dependence in simulation, where with more pile up, the efficiency gets lower. The uncertainties are too large in data to confirm this.

Further stability cross-checks with run number, magnet polarity and simulation versions show no significant dependency.

Figure 9.22: Dependence of electron reconstruction efficiency with respect to pile up, *i.e.* n_{PV}. Both samples in data and simulation have the full kinematic range. A slight dependence is seen in simulation. The uncertainties are too large in data to conform this.



9.9 Conclusion and use cases

The efficiency ratios obtained with this analysis will be a crucial input to measurements with electrons, especially for lepton-universality tests. The results of this analysis, albeit preliminary, have already been used to cross-check the systematic-uncertainty determination of the reconstruction efficiency of the 2019 update on R_{K^+} [17]. This measurement is one of the lepton-universality tensions discussed in chapter 1. The use of the efficiency ratios for this measurement will be discussed in section 9.9.1.

Aside from reconstruction efficiencies from electrons, this method can be applied to other modes and particle species as well. Some applications, together with a conclusion of the method, will be discussed in section 9.9.2.

9.9.1 Preliminary application of results: R_{K^+}

The measurement of the lepton-universality ratio of branching fractions R_{K^+}, defined for a specific dilepton invariant mass range by,

$$R_{K^+} = \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to K^+e^+e^-)} \ , \tag{9.8}$$

is sensitive to electron and muon reconstruction and selection differences.

To reduce this sensitivity, the experimental strategy of measuring R_{K^+} can be boiled down to using a double ratio, where the second ratio is both theoretically close to unity and it cancels out efficiency dependencies of electron and muons to first order. The relevant ratio, $r_{J/\psi}$, is known experimentally to high precision and is indeed very close to unity and given by,

$$r_{J/\psi} = \frac{\mathcal{B}(J/\psi \to \mu^+\mu^-)}{\mathcal{B}(J/\psi \to e^+e^-)} = 0.9984 \pm 0.0031 \ , \tag{9.9}$$

where the value is the PDG average [16]. Therefore, R_{K^+} can be measured with a reduction in systematic uncertainties by defining,

$$R_{K^+} = \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to K^+e^+e^-)} \frac{\mathcal{B}(B^+ \to J/\psi(\to e^+e^-)K^+)}{\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)} \simeq \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to K^+e^+e^-)} \ . \tag{9.10}$$

Table 9.11: Results of $r_{J/\psi}$, related to R_{K^+}, with two different reconstruction efficiency corrections. The reconstruction corrections clearly reduce the discrepancies of data and simulation, as the resulting ratios are closer to unity. Trigger categories eTOS, hTOS! and TIS! are related to triggering on the electrons, kaon or independent on signal respectively. Categories are mutually exclusive, by excluding previous categories. Values obtained from [120].

	eTOS	hTOS!	TIS!
Nominal	1.052 ± 0.005	1.053 ± 0.011	1.112 ± 0.009
With reconstruction correction	1.022 ± 0.005	1.017 ± 0.011	1.074 ± 0.008

Table 9.12: Systematic uncertainties of R_{K^+} related to reconstruction efficiency. The impact of reconstruction corrections, using the efficiency ratios of the presented analysis, is covered by the nominal assigned systematic uncertainty. Trigger categories eTOS, hTOS! and TIS! are related to triggering on the electrons, kaon or independent on signal respectively. Categories are mutually exclusive, by excluding previous categories. Values obtained from [120].

	eTOS	hTOS!	TIS!
Impact of reconstuction corrections	0.25%	0.03%	0.06%
Assigned reconstruction systematic	0.47%	0.02%	0.35%

An important pillar of the analysis is the measurement of $r_{J/\psi}$ and that its value is constant over the phase space of $B^+ \to J/\psi\, K^+$, indicating that electron versus muon efficiencies are sufficiently reproduced in simulation for this channel. The nominal electron reconstruction efficiency correction applied to simulation for this measurement is reweighing the distributions of the η of the electron to the one from data. The reasoning behind this is that the reconstruction efficiency is sensitive to material and the amount of material differs quite a bit for different ranges of η. Accordingly, a different reconstruction efficiency should alter the distribution of η. The results of $r_{J/\psi}$ for different trigger categories have been determined with the nominal correction and with the efficiency ratios obtained by the analysis presented in this thesis and are listed in table 9.11. It shows that indeed the efficiency ratios correct for some of the data versus simulation discrepancies, as $r_{J/\psi}$ gets closer to one.

On R_{K^+} itself, the effect is smaller and the difference in applying the efficiency ratio corrections or not is covered by the nominal systematic uncertainty determination, done with the η reweighing, shown in table 9.12.

9.9.2 Conclusion and further possibilities

The aforementioned application of the results already shows that data versus simulation discrepancies can be reduced. Accordingly, smaller systematic uncertainties arise. Hence, the method of this analysis will lead to analyses with electrons with smaller uncertainties. This is especially important for larger datasets of both Run 2 and the new LHCb detector in the future Run 3, as statistical uncertainties will go down. Note that so far just a part of the 2017 data set is used. The method can be applied, albeit with a lower efficiency, in the datasets of previous years through the use of the topological trigger, described in

section 3.1.3.

Aside from electrons, the method can be applied to muons as well, as the decay $J/\psi \to \mu^+\mu^-$ has practically the same branching fraction as $J/\psi \to e^+e^-$. This would be complementary to the already existing tracking efficiencies measured for muons [109]. This also allows the determination of the relative electron over muon efficiencies purely from data, which is of great interest for lepton-universality measurements. On top of the $B^+ \to J/\psi\, K^+$ mode, other b-hadron modes provide more data. Examples are $B^0 \to J/\psi\, K^{*0}$ and $B_s^0 \to J/\psi\, \phi$, which are implemented in the trigger of 2018 and are available through the topological trigger in previous years as well.

On top of measuring reconstruction efficiencies, one could use the sample to check efficiencies of selections in a later stage in the reconstruction and selection process.

To conclude, a new data-driven method has been presented in this chapter to measure differences in electron reconstruction efficiencies between data and simulation. The method has low systematic uncertainties and is so far dominated by statistical uncertainties. This will allow LHCb to enhance the accuracy in future measurements involving electrons, *e.g.* lepton-universality measurements.

Part IV

Appendices

Appendix A

Definitions of analytical PDFs

The definitions of uncommon analytical PDFs used throughout the thesis are listed in this section. Usually normalisations are not given, but most of them are analytically integrable.

A.1 Double-sided Crystal Ball

Adaptation from Crystal Ball function as originally developed for the Crystal Ball collaboration [121]. Extended to have on both sides a power-law tail. It is described by,

$$f^{\mathrm{DSCB}}(m) \propto \begin{cases} \exp\left(-\frac{(m-\mu)^2}{2\sigma^2}\right) & , \text{ if } -\alpha_1 < \frac{m-\mu}{\sigma} < \alpha_2 \\ A_1\left(\frac{n_1}{\alpha_1} - \alpha_1 - \frac{m-\mu}{\sigma}\right)^{-n_1} & , \text{ if } \frac{m-\mu}{\sigma} \leq -\alpha_1 \\ A_2\left(\frac{n_1}{\alpha_2} - \alpha_2 + \frac{m-\mu}{\sigma}\right)^{-n_2} & , \text{ if } \frac{m-\mu}{\sigma} \geq \alpha_2 \end{cases} , \tag{A.1}$$

where

$$A_1 = \left(\frac{n_1}{\alpha_1}\right)^{n_1} \exp\left(-\frac{\alpha_1^2}{2}\right) , \qquad A_2 = \left(\frac{n_2}{\alpha_2}\right)^{n_2} \exp\left(-\frac{\alpha_2^2}{2}\right) . \tag{A.2}$$

A.2 Hypatia

A generalisation of eq. (A.1) is the Hypatia function [66]. The concept is to generalise the width of the Gaussian to a distribution itself, to account for a spread in invariant-mass uncertainty. It is given by,

$$f^{\mathrm{Hypatia}}(m) \propto \begin{cases} \left((m-\mu)^2 + A_\lambda^2(\zeta)\sigma^2\right)^{\frac{1}{2}\lambda - \frac{1}{4}} e^{\beta(m-\mu)} \tilde{K}(m) & , \text{ if } -\alpha_1 < \frac{m-\mu}{\sigma} < \alpha_2 \\ \frac{G(\mu-\alpha_1\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{\left(1-m/(n\frac{G(\mu-\alpha_1\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{G'(\mu-\alpha_1\sigma,\mu,\sigma,\lambda,\zeta,\beta)}-\alpha_1\sigma)\right)^{n_1}} & , \text{ if } \frac{m-\mu}{\sigma} \leq -\alpha_1 \\ \frac{G(\mu-\alpha_2\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{\left(1-m/(n\frac{G(\mu-\alpha_2\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{G'(\mu-\alpha_2\sigma,\mu,\sigma,\lambda,\zeta,\beta)}-\alpha_2\sigma)\right)^{n_2}} & , \text{ if } \frac{m-\mu}{\sigma} \geq \alpha_2 \end{cases} , \tag{A.3}$$

where,

$$\tilde{K}(m) = K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{m-\mu}{A_\lambda(\zeta)\sigma}\right)^2}\right) , \tag{A.4}$$

G is the generalised hyperbolic distribution, K is a special Bessel function of the third kind and $A^2_\lambda(\zeta) = \frac{\zeta K_\lambda(\zeta)}{K_{\lambda+1}(\zeta)}$.

A.3 ARGUS

To describe continuous backgrounds, such as for partially-reconstructed decays, the ARGUS distribution is used [118]. It is defined by,

$$f^{\text{ARGUS}}(m) \propto m\sqrt{1-\left(\frac{m}{m_0}\right)^2}\exp\left[c\left(1-\left(\frac{m}{m_0}\right)^2\right)\right] \tag{A.5}$$

A.4 Thresholded double-sided Crystal Ball

In case of distribution with very long tails, which are limited by some threshold at very low values of m, a generalisation of eq. (A.1) is constructed by,

$$f^{\text{thDSCB}}(m) \propto \begin{cases} f^{\text{DSCB}}(m|\mu,\sigma,\alpha_1,\alpha_2,n_1,n_2) & , \text{ if } m \geq m_{\text{th}} \\ C\exp[c_0\log(m)+c_1\log(m)^2+c_2\log(m)^3] & , \text{ if } m < m_{\text{th}} \end{cases} \tag{A.6}$$

where

$$c_2 = \frac{m_{\text{th}}}{3\log(m_{\text{th}})^2}\left(\frac{c_0}{m_{\text{th}}}+\frac{2c_1\log(m_{\text{th}})}{m_{\text{th}}}-\frac{n_1}{\sigma\left(\frac{n_1}{\alpha_1}-\alpha_1-\frac{m-\mu}{\sigma}\right)}\right) ,$$

$$C = \frac{f^{\text{DSCB}}(m_{\text{th}})}{\exp[c_0\log(m_{\text{th}})+c_1\log(m_{\text{th}})^2+c_2\log(m_{\text{th}})^3]} ,$$

m_{th} is the theshold value of the mass, where the extra tail starts, c_0 and c_1 are shape parameters of the tail and the definition of c_2 ensures the derivative does not have a discontinuity.

A.5 Bernstein polynomials

For purpose of using polynomials as PDFs and subsequent likelihood maximization, it is convenient if the polynomial is positive for all values of its parameters. The Bernstein polynomials have this property and are defined by,

$$B^n(t|\vec{c}) = \sum_{i=0}^{n} c_i \frac{n!}{i!(n-i)!} t^i(1-t)^{n-i} , \tag{A.7}$$

where n is the order of the polynomial, $\vec{c}$ is a vector of parameters of length n and t is the variable in the range of 0 to 1, which can be used by any variable, x, in the range from x_{min} to x_{max}, by defining $t = (x - x_{\text{min}})/(x_{\text{max}} - x_{\text{min}})$ [117].

A.6 Misidentified Gaussian

The distribution of the invariant mass of a background with one misidentified particle with Gaussian resolution, defined also in eq. (4.6), is given by,

$$g^{misID}(m|\mu,\sigma,\vec{\lambda}) \propto \int_{x_{\min}}^{x_{\max}} dx \cdot g(m|\widetilde{\mu}(\mu,x),\sigma) \cdot f(x|\vec{\lambda}) \ , \tag{A.8}$$

where the Gaussian representing the invariant-mass PDF is given by,

$$g(m|\widetilde{\mu}(\mu,x),\sigma) = g(m,x|\mu,\sigma,\Delta m^2) = \exp\left(- \frac{(m - \sqrt{\mu^2 + \Delta m^2 x})^2}{2\sigma^2} \right) \tag{A.9}$$

and the distribution of x, the momentum fraction, is given by,

$$f(x|\vec{\lambda} = (x_{\min}, \Delta x, \vec{a}, \vec{b}, \vec{c})) \equiv \begin{cases} \sum_{i=0}^{3} b_i x^i & \text{for } x_{\min} \leq x < x_{\min} + \Delta x \\ \sum_{i=0}^{n} c_i e^{a_i x} & \text{for } x \geq x_{\min} + \Delta x \end{cases} \ , \tag{A.10}$$

where both the value and the derivatives, first and second, of the spline and the sum of exponentials at $x = x_{\min} + \Delta x$ are equal. The values of b_i are fixed. The sum of exponentials is constrained to approximate a power law by using,

$$x^{-k} \simeq \left(\sum_{i=0}^{N} p^{-ik} \exp\Big(- \frac{k}{p^i} \Big) \right)^{-1} \sum_{i=0}^{N} p^{-ik} \exp\Big(- \frac{k}{p^i} x \Big) \ , \tag{A.11}$$

where p is a scale parameter related to the range and distribution of x, set here to $p = 3$, and N is a precision parameter set to a compromise between precision and speed at $N = 5$ [70]. Therefore, the coefficients c_i and a_i are given by,

$$c_i = \left(\sum_{i=0}^{N} p^{-ik} \exp\Big(- \frac{k}{p^i} \Big) \right)^{-1} p^{-ik} \qquad\qquad a_i = -\frac{k}{p^i} \tag{A.12}$$

Implementations of eq. (A.8) are available at [122] for `RooFit` [73]. Integrals over the defined functions are all analytically solvable and are obtained using [123].

Appendix B

Selections of calibration and normalisation channels for part II

The goal of the selection of the calibration and normalisation channels for $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ is to be as close as possible to the signal selection, presented in chapter 3. The selections for the channels $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ are mostly the same as the ones used for a previous analysis, the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction [54]. The selections for these channels for that analysis are already presented in another thesis [57]. What is shown here are the differences with respect to the $B_s^0 \to \mu^+\mu^-$ analysis.

The differences for both the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ normalisation channels are only in the trigger selection, as the $B^0_{(d/s)} \to e^\pm\mu^\mp$ trigger has quite a different trigger strategy than the $B_s^0 \to \mu^+\mu^-$ analysis. For both the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ channels as normalisation, a pure `TOS` selection, *i.e.* trigger on signal, is chosen. The trigger selection for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ are shown in tables B.1 and B.2 respectively. All mentioned trigger lines are discussed in chapter 3.

Table B.1: Trigger selection for the normalisation channel $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$.

Candidate	Lines	Value	Note
μ^+ or μ^-	`L0Muon`	`TOS`	
μ^+, μ^- or K^+	`Hlt1TrackAllL0` or `Hlt1TrackMuon`	`TOS`	`Hlt1TrackMuon` only for $\mu^+\mu^-$
B^+ $[\mu^+\mu^-K^+]$	`Hlt2TopoMu2BodyBBDT` or `Hlt2TopoMu3BodyBBDT`	`TOS`	

Table B.2: Trigger selection for the normalisation channel $B^0 \to K^+\pi^-$.

Candidate	Lines	Value
K^+ or π^-	`L0Hadron`	`TOS`
K^+ or π^-	`Hlt1TrackAllL0`	`TOS`
B^0 $[K^+\pi^-]$	`Hlt2B2HH`	`TOS`

Appendix C

Electron-reconstruction efficiencies

This appendix contains supplementary information relevant to chapter 9.

C.1 Fits in kinematic bins

To highlight the region where most deviations are seen in the electron reconstruction efficiencies between data and simulation, the distributions of $m_{J/\psi K^{\pm}}$ in data including the PDF from the fit are shown for the pass and fail categories separately for the pseudo-rapidity bin of $1.9 \leq \eta < 2.9$ in the non-RF foil region in figs. C.1 to C.5.

Figure C.1: Distributions in data of the invariant mass, $m_{J/\psi K^{\pm}}$, of $B^+ \to J/\psi(\to e^+e^-)K^+$ decays with PDF resulting from the likelihood fit overlaid.

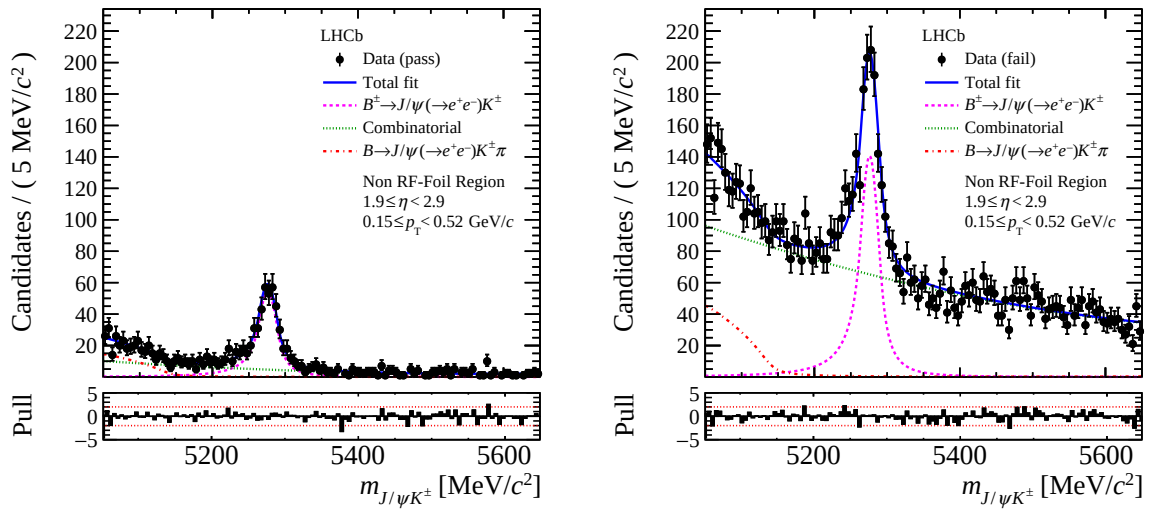


Figure C.2: Distributions in data of the invariant mass, $m_{J/\psi K^\pm}$, of $B^+ \to J/\psi(\to e^+e^-)K^+$ decays with PDF resulting from the likelihood fit overlaid.

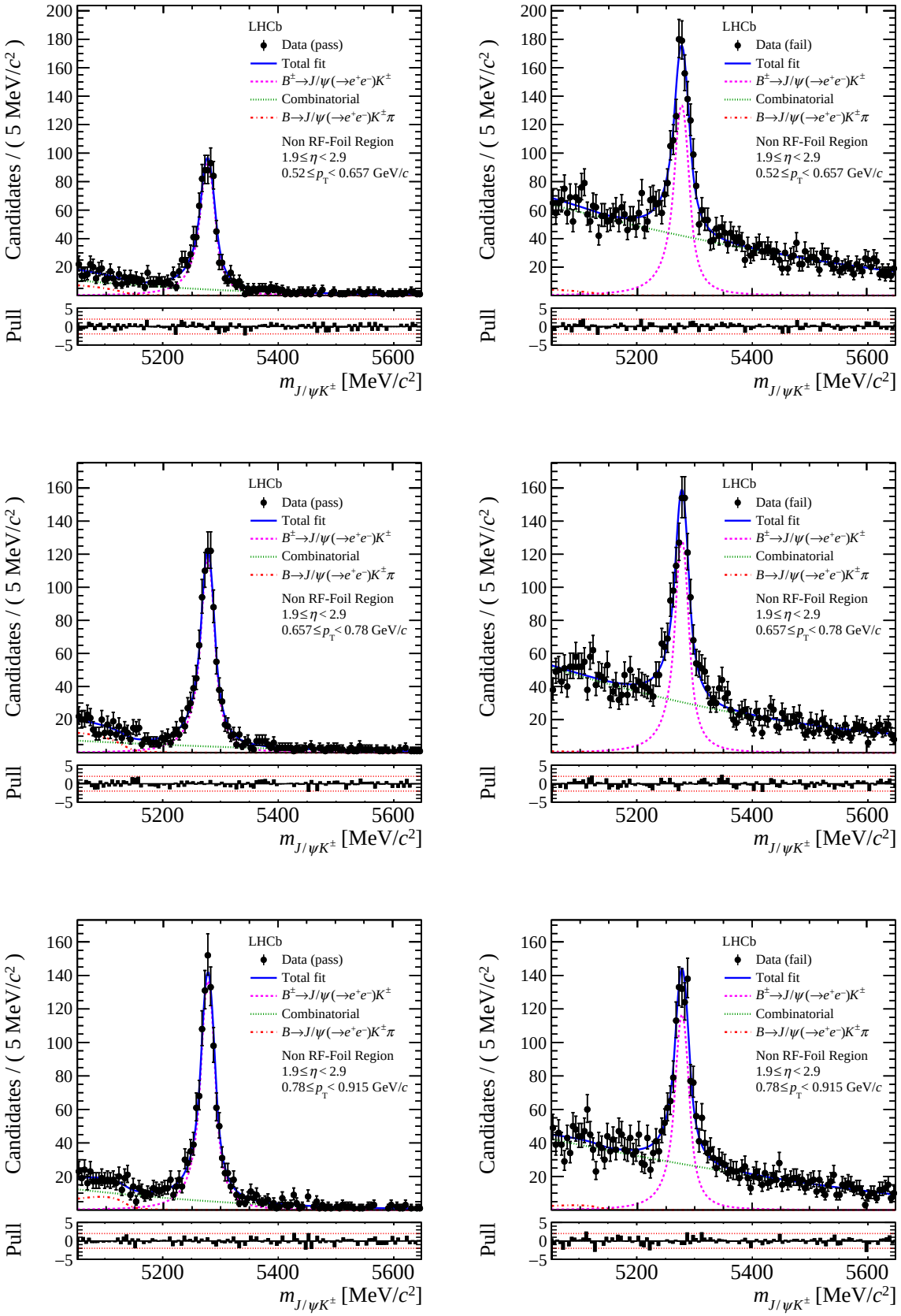


Figure C.3: Distributions in data of the invariant mass, $m_{J/\psi K^\pm}$, of $B^+ \to J/\psi(\to e^+e^-)K^+$ decays with PDF resulting from the likelihood fit overlaid.

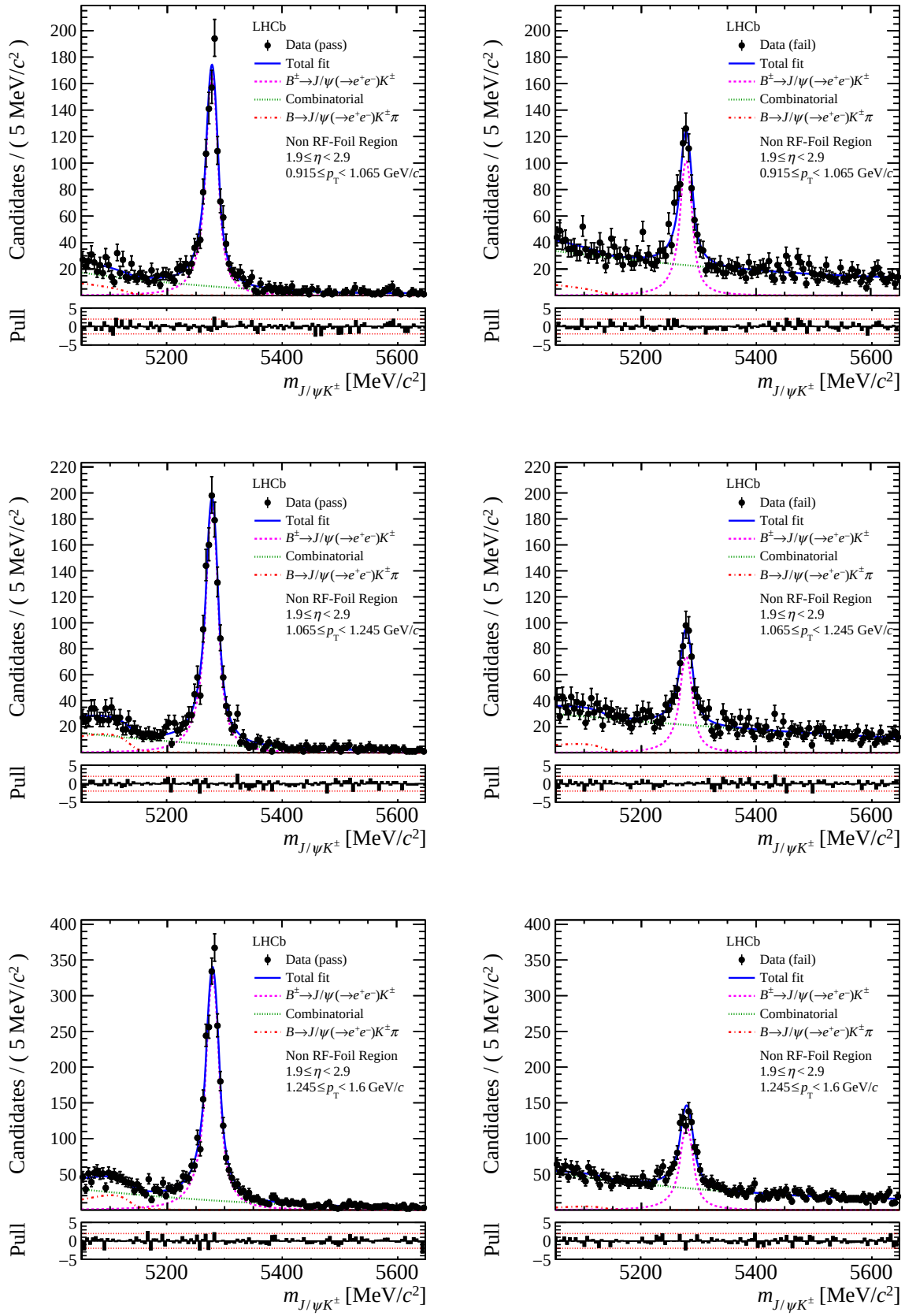


Figure C.4: Distributions in data of the invariant mass, $m_{J/\psi K^\pm}$, of $B^+ \to J/\psi(\to e^+e^-)K^+$ decays with PDF resulting from the likelihood fit overlaid.

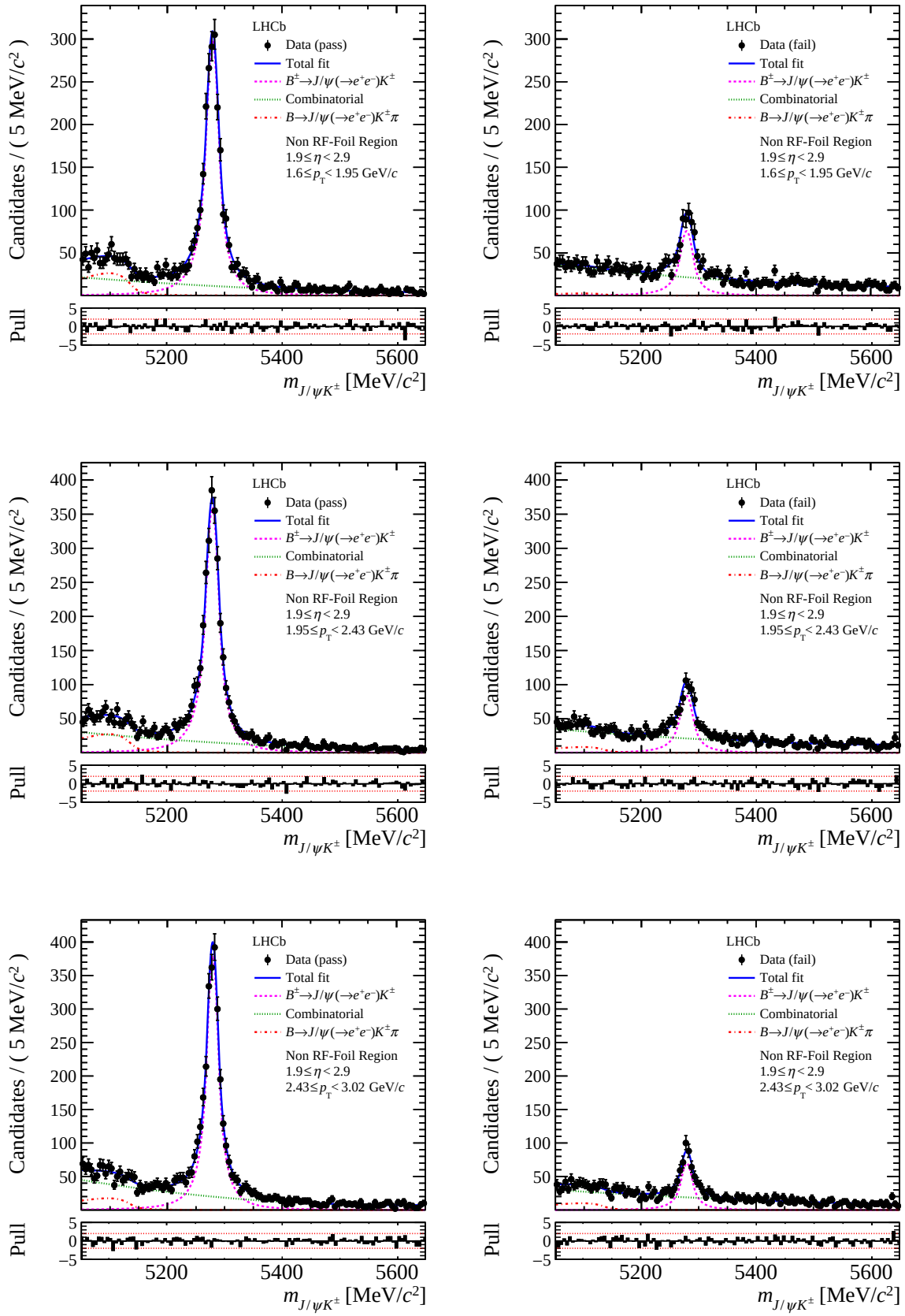
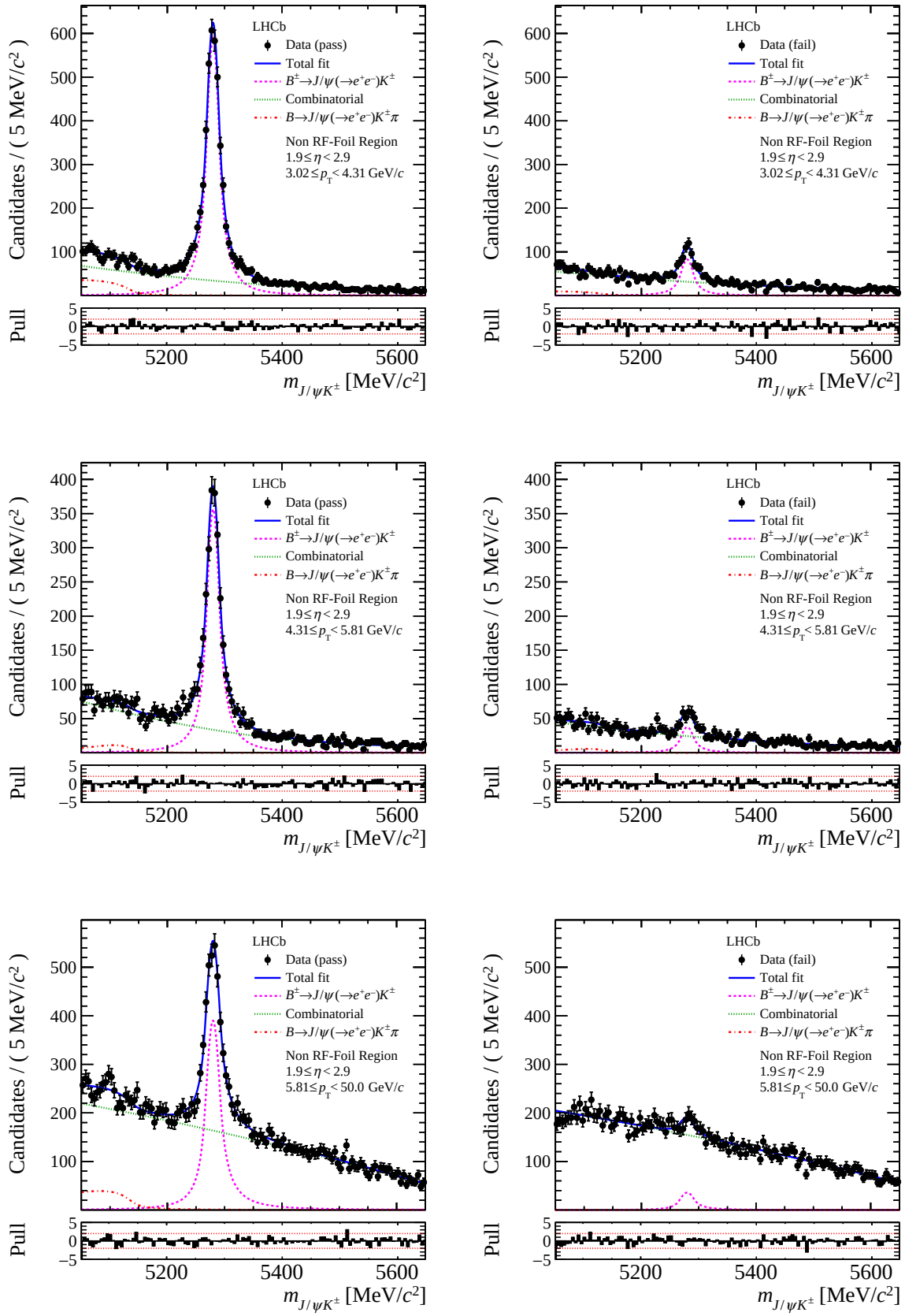


Figure C.5: Distributions in data of the invariant mass, $m_{J/\psi K^\pm}$, of $B^+ \to J/\psi(\to e^+e^-)K^+$ decays with PDF resulting from the likelihood fit overlaid.



C.2 Absolute efficiencies

The electron reconstruction efficiencies have as main result the ratio of efficiencies of data over simulation. The principle argument is that it reduces systematic uncertainties compared to absolute efficiencies. To quantify the argument, systematic biases due to the selection are determined in simulation. By linking the generator level particles to the simulated detector output, a selection can be based solely on truth matching reconstructed objects to the signal decays. By comparing this selection to the selection performed on data on top of the truth matching, the systematic biases can be assessed.

The resulting biases, per kinematic bin, are shown in figs. C.6 and C.7 for the non-RF-Foil and RF-Foil region respectively. It shows that they are of the order of a few percent and subsequently, to have a more precise result, the ratio of efficiencies is needed.

Figure C.6: Efficiencies in simulation per kinematic bin in the non-RF-Foil region after truth matching. One without additional selection, other including tag-and-probe selection. Discrepancies are of the order of few percent.

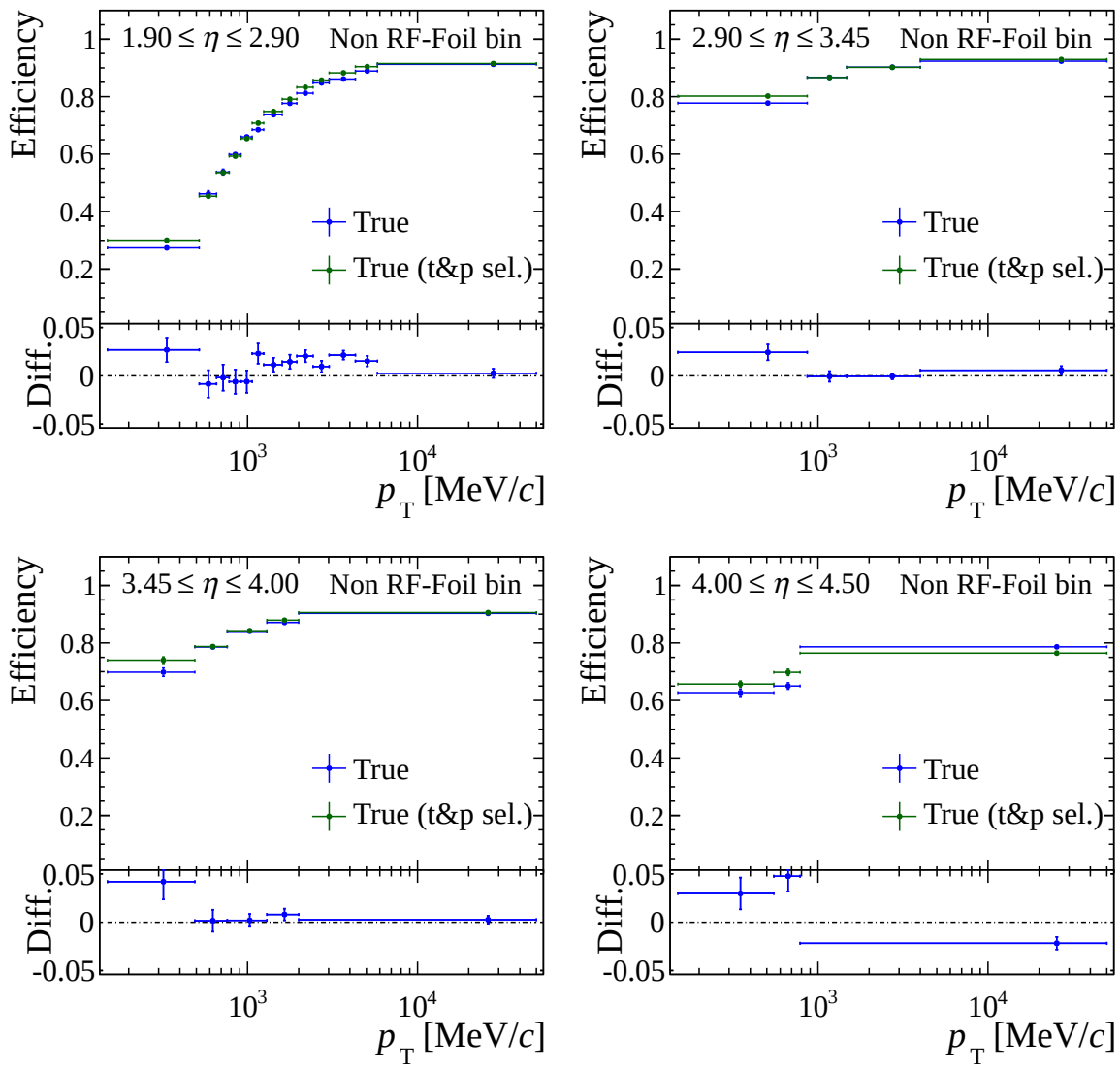
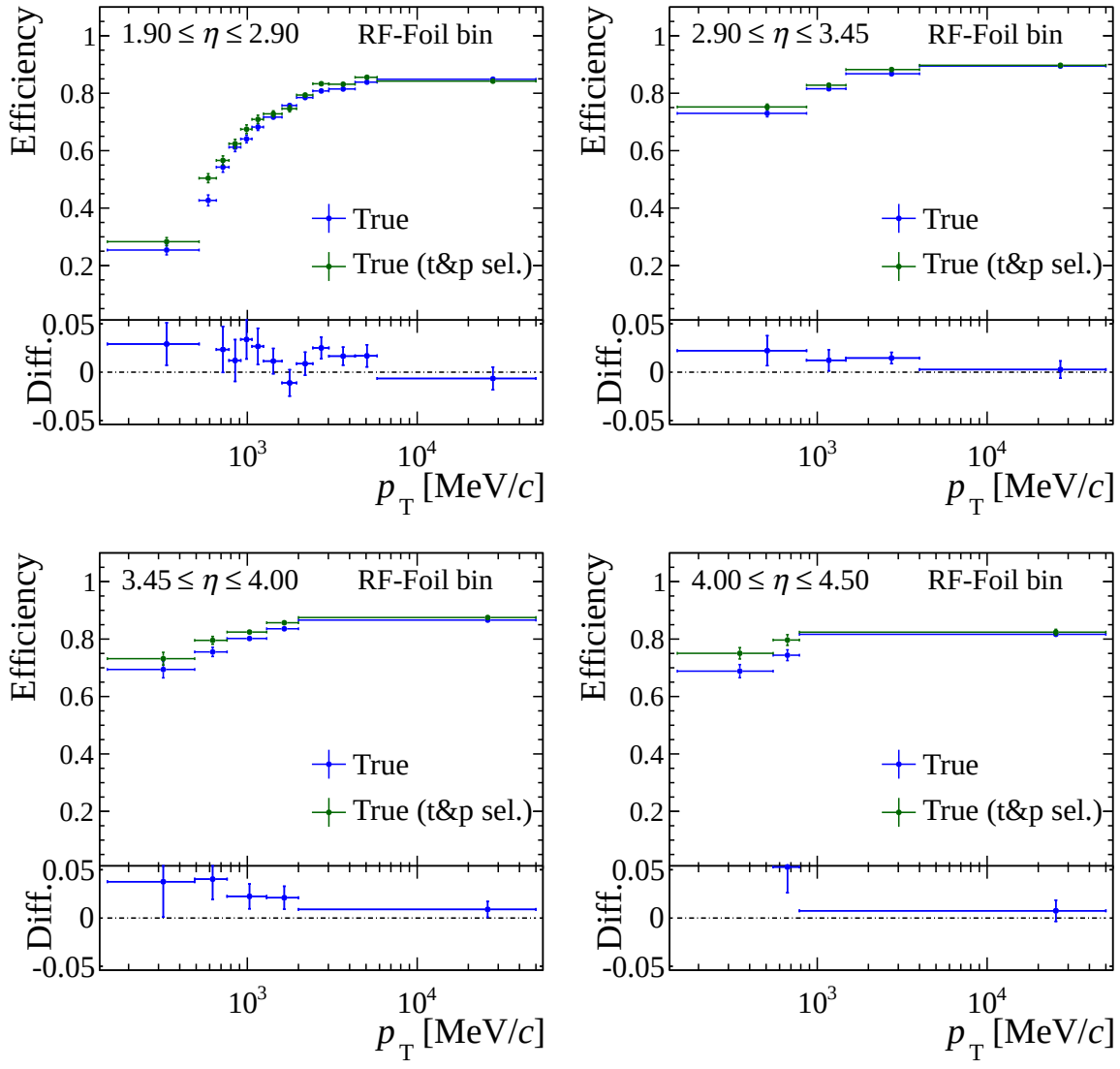


Figure C.7: Efficiencies in simulation per kinematic bin in the RF-Foil region after truth matching. One without additional selection, other including tag-and-probe selection. Discrepancies are of the order of few percent.



C.3 Fit shape dependencies

For the determination of the systematic uncertainties on the efficiency ratios due to the description of the $m_{J/\psi K^{\pm}}$ distribution for signal in the likelihood fit, an assessment in simulation has been made of pass versus fail dependencies of the parameters of the analytical PDFs that are used for the signal component.

The dependencies on the p_{T} of the probe of the differences between pass and fail for the width and mean of the core Gaussian of the double-sided Crystal Ball have been shown before in fig. 9.18, the dependencies of the average are shown in fig. C.8. The dependencies of the tail parameters on the p_{T} of the probe are shown in figs. C.9 and C.10.

No significant differences between the pass and fail categories are seen in these parameters.

Figure C.8: For simulation, the dependence on p_{T} of the mean (μ) and width (σ) of the Gaussian core of the double-sided Crystal Ball used for the signal PDF is shown.

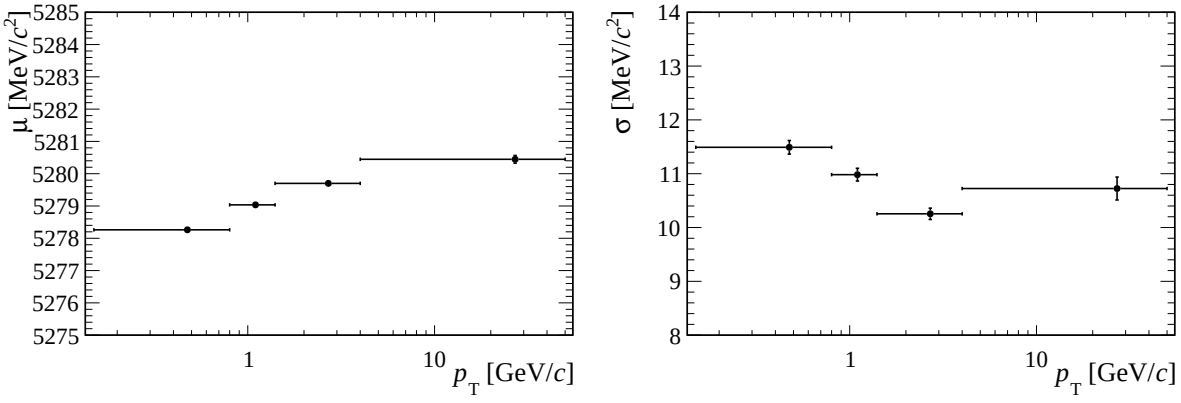


Figure C.9: For simulation, the dependence on p_{T} of the starting point, α, in terms of width of the Gaussian core of the double-sided Crystal Ball used for the signal PDF is shown. The left and right tail are denoted by the subscripts 1 and 2 respectively. They are determined separately for the pass and fail categories, shown right and left respectively. No significant differences are seen between pass and fail.

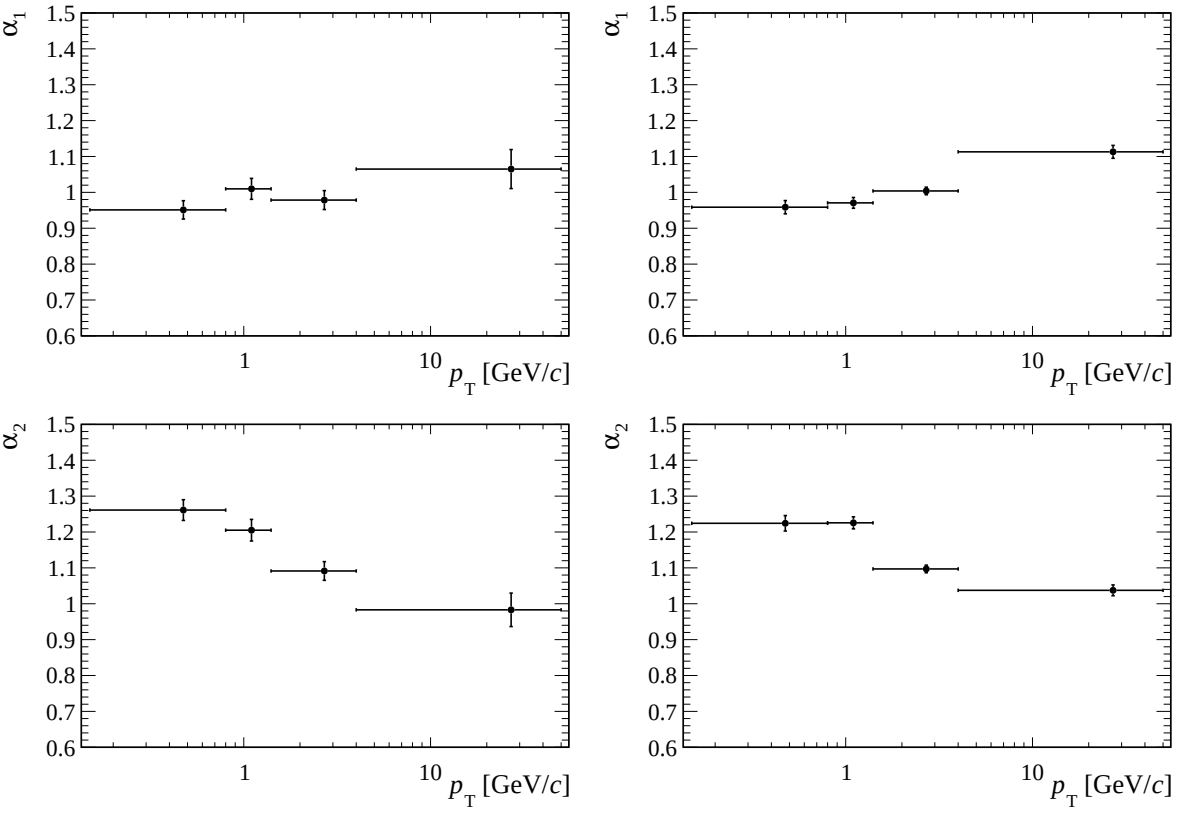
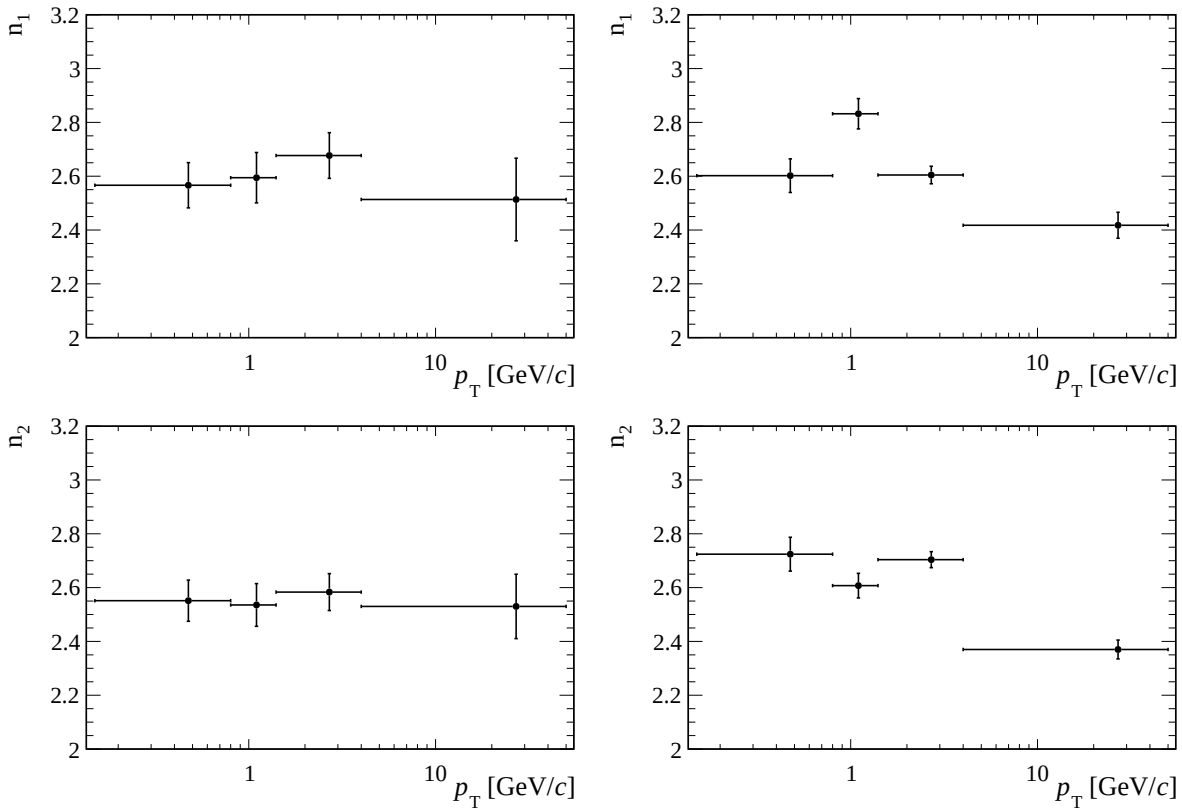


Figure C.10: For simulation, the dependence on p_{T} of the power of the tail, n, of the double-sided Crystal Ball used for the signal PDF is shown. The left and right tail are denoted by the subscripts 1 and 2 respectively. They are determined separately for the pass and fail categories, shown right and left respectively. No significant differences are seen between pass and fail.



References

[1] J. Woithe, G. J. Wiener, and F. F. Van der Veken, *Let's have a coffee with the Standard Model of particle physics!*, Physics Education **52** (2017) 034001.

[2] L. Baudis, *The search for dark matter*, European Review **26** (2018), no. 1 70–81.

[3] ATLAS, CMS, LHCb, E. Graverini, *Flavour anomalies: a review*, J. Phys. Conf. Ser. **1137** (2019), no. 1 012025, arXiv:1807.11373.

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[6] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[7] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44** (1972) 189.

[8] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA, 1995.

[9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, [157(1964)].

[10] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[11] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[12] ATLAS, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv:1207.7214.

[13] CMS, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv:1207.7235.

[14] MEG, A. M. Baldini *et al.*, *Search for the lepton flavour violating decay* $\mu^+ \to e^+\gamma$ *with the full dataset of the MEG experiment*, Eur. Phys. J. **C76** (2016), no. 8 434, arXiv:1605.05081.

[15] Super-Kamiokande, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003.

[16] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. D **98** (2018) 030001.

[17] LHCb, R. Aaij *et al.*, *Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **122** (2019), no. 19 191801, arXiv:1903.09252.

[18] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[19] LHCb, R. Aaij *et al.*, *Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays*, arXiv:1912.08139.

[20] S. L. Glashow, D. Guadagnoli, and K. Lane, *Lepton Flavor Violation in B Decays?*, Phys. Rev. Lett. **114** (2015) 091801, arXiv:1411.0565.

[21] D. Bečirević, S. Fajfer, N. Košnik, and O. Sumensari, *Leptoquark model to explain the B-physics anomalies, R_K and R_D*, Phys. Rev. **D94** (2016), no. 11 115021, arXiv:1608.08501.

[22] A. Crivellin, D. Müller, and T. Ota, *Simultaneous explanation of $R(D^{(\star)})$ and $b \to s\mu^+\mu^-$: the last scalar leptoquarks standing*, JHEP **09** (2017) 040, arXiv:1703.09226.

[23] S. M. Bilenky, S. T. Petcov, and B. Pontecorvo, *Lepton Mixing, $\mu \to e + \gamma$ Decay and Neutrino Oscillations*, Phys. Lett. **67B** (1977) 309.

[24] I. Esteban *et al.*, *Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ_{23}, $\delta_{\rm CP}$, and the mass ordering*, JHEP **01** (2019) 106, arXiv:1811.05487.

[25] A. Crivellin *et al.*, *Lepton-flavour violating B decays in generic Z' models*, Phys. Rev. **D92** (2015) 054013, arXiv:1504.07928.

[26] I. de Medeiros Varzielas and G. Hiller, *Clues for flavor from rare lepton and quark decays*, JHEP **06** (2015) 072, arXiv:1503.01084.

[27] A. Ilakovac, *Lepton flavor violation in the standard model extended by heavy singlet Dirac neutrinos*, Phys. Rev. **D62** (2000) 036010, arXiv:hep-ph/9910213.

[28] R. A. Diaz, R. Martinez, and C. E. Sandoval, *Improving bounds on flavor changing vertices in the two Higgs doublet model from $B^0 - \overline{B}^0$ mixing*, Eur. Phys. J. **C46** (2006) 403, arXiv:hep-ph/0509194.

[29] J. C. Pati and A. Salam, *Lepton Number as the Fourth Color*, Phys. Rev. **D10** (1974) 275, Erratum-ibid. **D11** (1975) 703.

[30] D. Bečirević, O. Sumensari, and R. Zukanovich Funchal, *Lepton flavor violation in exclusive $b \to s$ decays*, Eur. Phys. J. **C76** (2016), no. 3 134, arXiv:1602.00881.

[31] LHCb, R. Aaij *et al.*, *Search for the lepton-flavor violating decays $B_s^0 \to e^\pm\mu^\mp$ and $B^0 \to e^\pm\mu^\mp$*, Phys. Rev. Lett. **111** (2013) 141801, arXiv:1307.4889.

[32] K. De Bruyn *et al.*, *Branching Ratio Measurements of B_s Decays*, Phys. Rev. **D86** (2012) 014027, arXiv:1204.1735.

[33] Y. Ohnishi *et al.*, *Accelerator design at SuperKEKB*, Progress of Theoretical and Experimental Physics **2013** (2013) , 03A011.

[34] Belle-II, T. Abe *et al.*, *Belle II Technical Design Report*, arXiv:1011.0352.

[35] L. Evans and P. Bryant, *LHC machine*, Journal of Instrumentation **3** (2008) S08001.

[36] LHCb, R. Antunes-Nobrega *et al.*, *LHCb reoptimized detector design and performance: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2003.

[37] LHCb, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[38] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[39] Yu. Guz, *The LHCb Calorimeter system: design, performance and upgrade*, JINST **12** (2017), no. 07 C07024.

[40] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[41] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[42] LHCb, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019), no. 04 P04013, arXiv:1812.10790.

[43] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[44] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, Journal of Physics: Conference Series **331** (2011) 032047.

[45] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001), no. 1 152 , BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[46] E. Barberio and Z. Wąs, *PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0*, Computer Physics Communications **79** (1994), no. 2 291 .

[47] P. Golonka and Z. Wąs, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv:hep-ph/0506026.

[48] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[49] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[50] LHCb, R. Aaij *et al.*, *Search for the lepton-flavour violating decays* $B^0_{(s)} \to e^{\pm}\mu^{\mp}$, JHEP **03** (2018) 078, arXiv:1710.04111.

[51] M. Williams *et al.*, *The HLT2 Topological Lines*, Tech. Rep. LHCb-PUB-2011-002, CERN, Geneva, Jan, 2011.

[52] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, Tech. Rep. LHCb-PUB-2011-016, CERN, Geneva, Sep, 2011. LHCb-PUB-2011-016.

[53] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, arXiv:1210.6861.

[54] LHCb, R. Aaij *et al.*, *Measurement of the* $B^0_s \to \mu^+\mu^-$ *branching fraction and effective lifetime and search for* $B^0 \to \mu^+\mu^-$ *decays*, Phys. Rev. Lett. **118** (2017), no. 19 191801, arXiv:1703.05747.

[55] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** (1933) 289.

[56] A. Hoecker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039.

[57] S. Tolk, *Discovery of Rare B Decays*, PhD thesis, Groningen U., 2016.

[58] LHCb, R. Aaij *et al.*, *Measurements of the* $\Lambda^0_b \to J/\psi\Lambda$ *decay amplitudes and the* Λ^0_b *polarisation in pp collisions at* $\sqrt{s} = 7$ *TeV*, Phys. Lett. **B724** (2013) 27, arXiv:1302.5578.

[59] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10 100001.

[60] LHCb, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio* f_s/f_d *and its dependence on B meson kinematics*, JHEP **04** (2013) 001, arXiv:1301.5286, f_s/f_d value updated in LHCb-CONF-2013-011.

[61] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, Tech. Rep. LHCb-PUB-2014-039, CERN, Geneva, May, 2014.

[62] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.

[63] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[64] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[65] R. Barlow, *Extended maximum likelihood*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **297** (1990), no. 3 496 .

[66] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, arXiv:1312.5000.

[67] T. M. Karbach and M. Schlupp, *Constraints on Yield Parameters in Extended Maximum Likelihood Fits*, arXiv:1210.7141.

[68] M. Artuso *et al.*, *Measurement of the CP violating asymmetry a^s_{sl}*, Tech. Rep. LHCb-ANA-2016-004, CERN, May, 2016. Linked to LHCb-PAPER-2016-013.

[69] S. Bachman *et al.*, *Measurement of $d^0 - \bar{D}^0$ oscillations with 'wrong-sign' $d^{*+} \to d^0(\to k^+\pi^-)\pi^+$ decays*, Tech. Rep. LHCb-ANA-2012-088, CERN, Jan, 2012. Linked to LHCb-PAPER-2012-038.

[70] T. Bochud and D. Challet, *Optimal approximations of power-laws with exponentials*, arXiv:physics/0605149.

[71] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198, arXiv:hep-ex/0011057.

[72] F. James and M. Roos, *MINUIT-a system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun. **10** (1975), no. 6 343.

[73] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, arXiv:physics/0306116.

[74] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60.

[75] A. L. Read, *Presentation of search results: the CL_s technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002) 2693.

[76] K. De Bruyn *et al.*, *Probing New Physics via the $B^0_s \to \mu^+\mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737.

[77] L. Moneta, K. Cranmer, G. Schott, and W. Verkerke, *The RooStats project*, arXiv:1009.1003.

[78] J. Kumar and D. London, *New physics in $b \to se^+e^-$?*, Phys. Rev. **D99** (2019), no. 7 073008, arXiv:1901.04516.

[79] B. Fornal, S. A. Gadam, and B. Grinstein, *Left-Right SU(4) Vector Leptoquark Model for Flavor Anomalies*, Phys. Rev. **D99** (2019), no. 5 055025, arXiv:1812.01603.

[80] A. D. Smirnov, *Vector leptoquark mass limits and branching ratios of $K^0_L, B^0, B_s \to l^+_i l^-_j$ decays with account of fermion mixing in leptoquark currents*, Mod. Phys. Lett. **A33** (2018) 1850019, arXiv:1801.02895.

[81] A. V. Povarov, *Constraints on Vector Leptoquark Masses from Data on the $K_L^0 \to l_i^+ l_j^-$ and $B^0, B_s^0 \to l_i^+ l_j^-$ Decays*, Phys. Part. Nucl. Lett. **16** (2019), no. 1 1.

[82] D. E. Hazard and A. A. Petrov, *Radiative lepton flavor violating B, D, and K decays*, Phys. Rev. **D98** (2018), no. 1 015027, arXiv:1711.05314.

[83] J. Aebischer, J. Kumar, P. Stangl, and D. M. Straub, *A Global Likelihood for Precision Constraints and Flavour Anomalies*, Eur. Phys. J. **C79** (2019), no. 6 509, arXiv:1810.07698.

[84] S. Hu, S. M.-Y. Wong, and F. Xu, *Probing Sterile Neutrino via Lepton Flavor Violating Decays of Mesons*, arXiv:1904.00568.

[85] K.-S. Sun and X.-Y. Yang, *The LFV decays $B_{d,s}^0 \to e\mu(e\tau, \mu\tau)$ with one neutral singlet scalar*, Mod. Phys. Lett. **A33** (2018), no. 36 1850214, arXiv:1901.03805.

[86] J.-J. Zhang, M. He, X.-G. He, and X.-B. Yuan, *Flavor Violating Higgs Couplings in Minimal Flavor Violation*, JHEP **02** (2019) 007, arXiv:1807.00921.

[87] LHCb, R. Aaij *et al.*, *Measurement of the $B^\pm$ production cross-section in pp collisions at $\sqrt{s} = 7$ and 13 TeV*, JHEP **12** (2017) 026, arXiv:1710.04921.

[88] LHCb collaboration, I. Bediaga *et al.*, *Framework TDR for the LHCb Upgrade: Technical Design Report*, Tech. Rep. CERN-LHCC-2012-007. LHCb-TDR-12, Apr, 2012.

[89] L. E. The LHCb Collaboration, *Physics case for an LHCb Upgrade II*, Tech. Rep. LHCb-PUB-2018-009. CERN-LHCb-PUB-2018-009, CERN, Geneva, May, 2018.

[90] E. Barberio, B. van Eijk, and Z. Wąs, *Photos — a universal monte carlo for qed radiative corrections in decays*, Computer Physics Communications **66** (1991), no. 1 115 .

[91] J. Van Tilburg and M. Merk, *Track simulation and reconstruction in LHCb*, 2005. CERN-THESIS-2005-040.

[92] H. Bethe and W. Heitler, *On the Stopping of fast particles and on the creation of positive electrons*, Proc. Roy. Soc. Lond. **A146** (1934) 83.

[93] Y.-S. Tsai, *Pair production and bremsstrahlung of charged leptons*, Rev. Mod. Phys. **46** (1974) 815.

[94] L. D. Landau and I. J. Pomeranchuk, *The limits of applicability of the theory of Bremsstrahlung by electrons and of the creation of pairs at large energies*, Dokl. Akad. Nauk. SSSR **92** (1953) 535.

[95] L. D. Landau and I. J. Pomeranchuk, *Electron-cascade processes at ultra-high energies*, Dokl. Akad. Nauk. SSSR **92** (1953) 735.

[96] D. J. Griffiths, *Introduction to electrodynamics; 4th ed.*, Pearson, Boston, MA, 2013. Re-published by Cambridge University Press in 2017, doi: 10.1017/9781108333511.

[97] H. W. Koch and J. W. Motz, *Bremsstrahlung cross-section formulas and related data*, Rev. Mod. Phys. **31** (1959) 920.

[98] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, Tech. Rep. LHCb-PUB-2011-001, CERN, Geneva, Jan, 2011.

[99] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME–Journal of Basic Engineering **82** (1960), no. Series D 35.

[100] E. Bos, *Reconstruction of charged particles in the LHCb experiment*, 2010. Presented on 10 Feb 2010.

[101] H. E. Rauch, C. T. Striebel, and F. Tung, *Maximum likelihood estimates of linear dynamic systems*, AIAA Journal **3** (1965), no. 8 1445, arXiv:https://doi.org/10.2514/3.3166.

[102] R. Frühwirth, *Track fitting with non-gaussian noise*, Computer Physics Communications **100** (1997), no. 1 1 .

[103] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*, .

[104] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, Tech. Rep. ATLAS-CONF-2012-047, CERN, Geneva, May, 2012.

[105] O. Deschamps *et al.*, *Photon and neutral pion reconstruction*, Tech. Rep. LHCb-2003-091, CERN, Geneva, Sep, 2003.

[106] V. Breton, N. Brun, and P. Perret, *A clustering algorithm for the LHCb electromagnetic calorimeter using a cellular automaton*, Tech. Rep. LHCb-2001-123, CERN, Geneva, Sep, 2001.

[107] I. Belyaev, V. Egorychev, D. Golubkov, and O. Stenyakin, *The X-correction for the e^+ and e^- track-cluster matching procedure.*, Tech. Rep. LHCb-INT-2014-029, CERN, Geneva, Jun, 2014.

[108] H. Terrier and I. Belyaev, *Particle identification with LHCb calorimeters*, Tech. Rep. LHCb-2003-092, CERN, Geneva, Sep, 2003.

[109] LHCb, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015), no. 02 P02007, arXiv:1408.1251.

[110] LHCb, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, Journal of Instrumentation **14** (2019) P11023, arXiv:1909.02957.

[111] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[112] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, *ReDecay: A novel approach to speed up the simulation at LHCb*, Eur. Phys. J. **C78** (2018), no. 12 1009, arXiv:1810.10362.

[113] F. Gao and L. Han, *Implementing the Nelder-Mead simplex algorithm with adaptive parameters*, Computational Optimization and Applications **51** (2012) 259.

[114] R. Aaij *et al.*, *Tesla : an application for real-time data analysis in High Energy Physics*, Comput. Phys. Commun. **208** (2016) 35, arXiv:1604.05596.

[115] R. Aaij *et al.*, *A comprehensive real-time analysis model at the LHCb experiment*, Journal of Instrumentation **14** (2019) P04006, arXiv:1903.01360.

[116] LHCb, R. Aaij *et al.*, *Observation of CP Violation in Charm Decays*, Phys. Rev. Lett. **122** (2019), no. 21 211803, arXiv:1903.08726.

[117] G. G. Lorentz, *Bernstein polynomials*, Univ. Toronto Press, 1953.

[118] H. Albrecht *et al.*, *Search for hadronic b–u decays*, Physics Letters B **241** (1990) 278.

[119] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[120] P. Alvarez Cartelle, T. Humair, and M. Patel, *Test of lepton flavour universality using $B^+ \to K^+ l^+ l^-$ decays*, Tech. Rep. LHCb-ANA-2017-042, CERN, Geneva, Jun, 2017. related to [17].

[121] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis, Cracow, INP, 1986.

[122] M. van Veghel, *RooMisID*, https://github.com/MaartenvanV/RooMisID, 2016.

[123] W. R. Inc. *Mathematica, Version 10.3*, October, 2015.

Summary

In the course of the 20th century, physicists developed a very successful theory, called the Standard Model, describing all known fundamental particles and their interactions, aside from gravity, down to very small length scales and up to high energies. Aside from its success at these levels, extrapolations towards cosmological length and time scales reveal its shortcomings. One of the most prominent issues is that the matter content of the Universe cannot be accounted for based on the Standard Model, neither its total amount nor the balance between matter and antimatter.

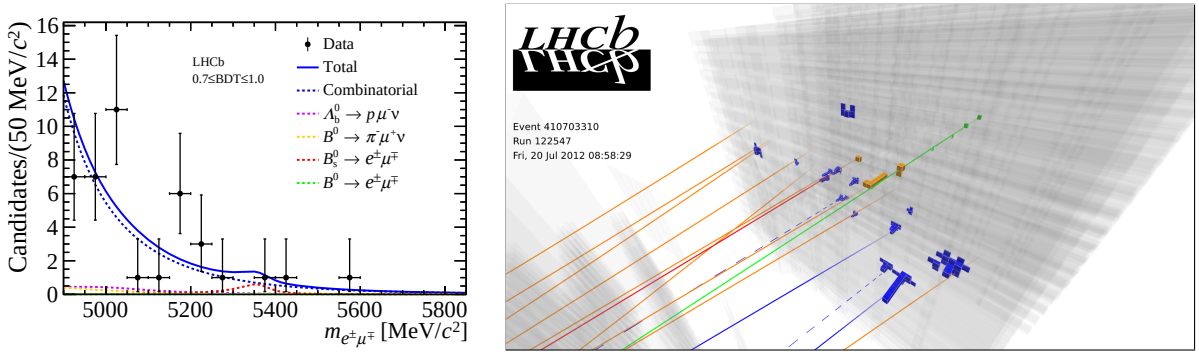
To solve these and other remaining issues, physicists are exploring higher energy scales to see if the Standard Model breaks and subsequently hints towards more fundamental laws of Nature. One approach is to indirectly probe high energy scales with precision measurements of processes at relatively lower energy scales. This is possible because virtual particles can alter processes at energy scales lower than their mass within a time scale limited by the uncertainty principle, although their effect is limited by the energy scale of the process relative to the virtual particle's mass. Hence, precision measurements are required.

Once such approach is the study of b-hadrons, which are particles containing at least a beauty quark, which is one of the heaviest fundamental particles of the Standard Model. Many decay processes are kinematically possible due to its high mass and subsequently many processes can be studied. In addition, the heavier particles of the Standard Model are less constrained by measurements than the lighter ones. One of the main sources of these b-hadrons are high energy proton-proton collisions, like the ones from the particle accelerator LHC at CERN in Geneva. There, more than a trillion a year are produced, allowing for the study of b-hadron decays that are rare or forbidden according to the Standard Model with great precision. While on an absolute scale hypothetical contributions to these decays from physics at higher energy scales can be tiny, their effect can be relatively large compared to the contribution from Standard Model particles. Hence, searches for these decays are effectively precision measurements.

In the last decade, hints of deviations have appeared in measurements of a property of the Standard Model called lepton universality, which essentially equates to that electrons, muons and taus interact in the same way with force mediators. A closely related property is lepton-flavour conservation (the conservation of lepton-family number). Hypothetical decays that violate lepton-flavour conservation are therefore forbidden in the Standard Model and provide an excellent testing ground and help to constrain models aimed at providing explanations for deviations from lepton universality.

The first part of this thesis is dedicated to a search for the lepton-flavour violating decays $B^0 \to e^{\pm}\mu^{\mp}$ and $B_s^0 \to e^{\pm}\mu^{\mp}$. This search is performed with data obtained with the LHCb detector in 2011 and 2012 at the LHC at CERN in Geneva. The LHCb detector is

Figure S.1: On the left: invariant-mass distribution of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ candidates with a very high signal-like classification score. On the right: illustration of partial detector output of the event containing the most-signal like $B_s^0 \to e^\pm\mu^\mp$ candidate, as shown in the distribution on the left. Details on this illustration are given in section 5.4.



specifically designed to measure properties of b-hadron decays. It accurately tracks the path of decay products traveling through the detector to obtain their original momentum, point of origin and particle type.

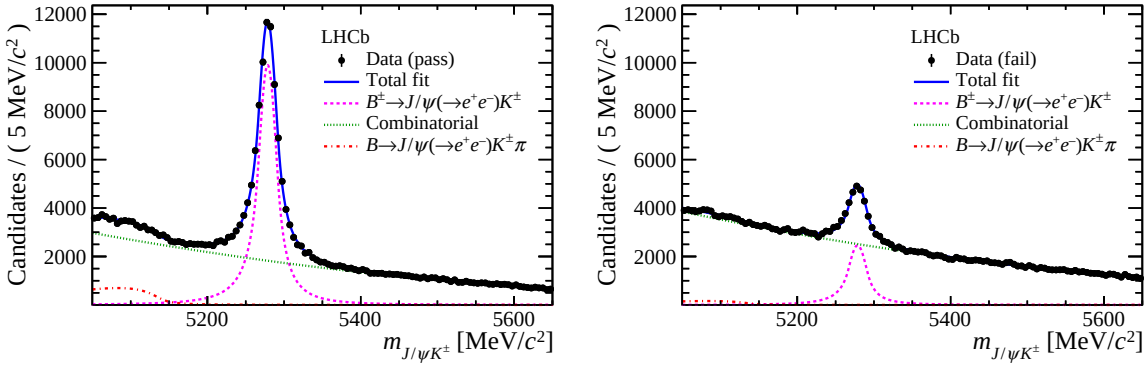
Candidates for the decays of $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ are built with combinations of an electron and a muon that are reconstructed in the detector and subsequently classified and selected with the use of machine learning based on variables constructed with the detector output. Then, the hypothetical contribution of the $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ decays are obtained from the distribution of the invariant mass of the combination of the electron and muon, which should peak at the B^0 or B_s^0 mass respectively. These spectra are divided into levels of the classifier for optimal separation of signal and background. Large part of the work goes into calibrating the detector response, making both use of simulation and data. For the latter, well-known decay channels like $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ are used as control channels and to turn yields obtained from the invariant-mass distributions into branching fractions, *i.e.* the probability for a decay to go to a certain final state.

The invariant-mass distribution of candidates with a very high classification score is shown in fig. S.1. It shows no peaks at the B^0 and B_s^0 mass of about 5280 MeV/c^2 and 5367 MeV/c^2 respectively. Hence, this analysis did not see a significant contribution of the $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ decays. Consequently, upper limits on the branching fractions have been set at 1.3×10^{-9} and 6.3×10^{-9} at 90% confidence level respectively. This makes it the world's best limit, improving the previous limit of LHCb by two to three times.

The last part of the thesis covers a study of the electron-reconstruction performance of the LHCb detector. The quantification of this performance is of essential importance to reduce and determine systematic uncertainties for *e.g.* lepton-universality measurements with the LHCb detector. This study is the first data-driven method for the determination of reconstruction efficiencies dedicated to electrons at LHCb.

The electron stands out as it is has a much smaller mass than other charged particles, which causes it to loose much more energy in the form of bremsstrahlung while traveling through the detector material. This makes the electron reconstruction more sensitive to the detector description in simulation and less efficient, and the electron momentum more difficult to infer.

Figure S.2: Distributions of the invariant mass of $B^+ \to J/\psi(\to e^+e^-)K^+$ candidates for fully-reconstructed (left) and partially-reconstructed (right) probe electrons. It shows that decays of $B^+ \to J/\psi(\to e^+e^-)K^+$ can be reconstructed with limited knowledge of the detector on one of the electrons, hence efficiencies of the full reconstruction can be measured.



To measure the reconstruction efficiency, the decay of $B^+ \to J/\psi(\to e^+e^-)K^+$ is exploited, where the probe electron is found with a minimal amount of information from the detector. For this purpose, only information from the vertex detector is used. The rest of the decay products are fully reconstructed and additional strict requirements are set to reduce backgrounds. The kinematics of this decay allows to determine the momentum of the probe electron and subsequently the full invariant-mass of the $e^+e^-K^+$ combination can be calculated. The efficiency can then be obtained by determining how many times the probe electron can be fully reconstructed or not. Distributions of the invariant-mass of $B^+ \to J/\psi(\to e^+e^-)K^+$ candidates are shown for fully-reconstructed (pass) and partially-reconstructed (fail) electron tracks in fig. S.2. It shows that the full decay can be reconstructed without the complete detector information on one of the electrons.

The results of the study have so far been used to cross-check systematic uncertainties for one of the lepton-universality measurements. For the future, this method will likely be the standard.

Samenvatting

In de loop van de 20e eeuw hebben fysici een erg succesvolle theorie ontwikkeld, genaamd het Standaard Model, die alle bekende fundamentele deeltjes en interacties beschrijft, behalve de zwaartekracht, tot erg kleine lengteschalen en tot hoge energieën. Ondanks haar succes op deze schalen, ontmaskeren extrapolaties naar kosmologische lengte- en tijdschalen haar tekortkomingen. Eén van de meest prominente problemen is dat de hoeveelheid materie in het Universum niet verklaard kan worden door het Standaard Model, zowel in termen van totale hoeveelheid als balans tussen materie en antimaterie.

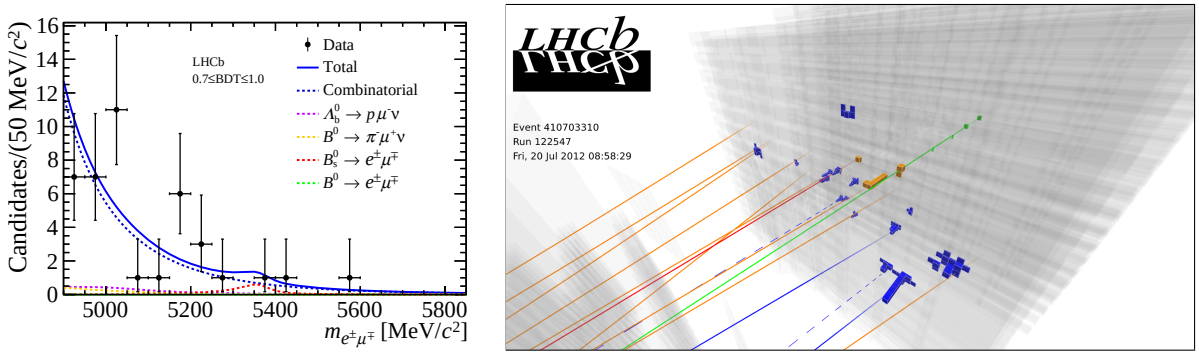
Om deze en andere problemen op te lossen onderzoeken fysici hogere energieschalen om te zien of het Standaard Model breekt en vervolgens hints geeft naar meer fundamentelere beschrijvingen van de Natuur. Eén benadering is het indirect peilen van hogere energieschalen met precisiemetingen van processen op relatief lage energie. Dit is mogelijk omdat virtuele deeltjes processen kunnen beïnvloeden op relatief lagere energieschalen dan hun massa binnen een tijdschaal gedicteerd door het onzekerheidsprincipe, alhoewel het effect wordt onderdrukt door de energie van het proces relatief aan de massa van het virtuele deeltje. Vandaar dat precisiemetingen nodig zijn.

De studie van b-hadrons is zo'n benadering. Deze b-hadrons zijn deeltjes die op zijn minst een beauty quark bevatten, welke één van de zwaarste fundamentele deeltjes is van het Standaard Model. Veel vervalsprocessen zijn mogelijk door deze hoge massa en vervolgens kunnen veel processen kunnen worden bestudeerd. Bovendien zijn de eigenschappen van zwaardere deeltjes van het Standaard Model minder nauwkeurig gemeten door al bestaande metingen dan de lichtere. Eén van de hoofdbronnen van deze b-hadrons zijn protonbotsingen, zoals diegenen van de deeltjesversneller LHC op CERN in Genève. Daar worden meer dan een biljoen per jaar geproduceerd, wat het bestuderen van vervallen van b-hadrons die zeldzaam of verboden zijn volgens het Standaard Model mogelijk maakt tot grote precisie. Terwijl op absolute schaal hypothetische bijdragen aan deze vervallen van fysica op hogere energieschalen klein kunnen zijn, kan hun effect relatief groot zijn ten opzichte van de bijdrage van deeltjes van het Standaard Model. Vandaar dat zoektochten naar deze vervallen effectief gezien precisiemetingen zijn.

In het laatste decennium zijn hints van afwijkingen opgedoken in metingen aan een eigenschap van het Standaard Model genaamd leptonuniversaliteit (elektronen, muonen en taus interacteren op dezelfde wijze met ijkbosonen). Een nauw verwante eigenschap is leptonsmaakbehoud (behoud van leptonfamiliegetal). Hypothetische vervallen die leptonsmaakbehoud schenden zijn daarom verboden in het Standaard Model en verschaffen een voortreffelijk proefterrein en helpen het beperken van modellen die potentieel deze afwijkingen kunnen verklaren.

Het eerste gedeelte van deze dissertatie is toegewijd aan een zoektocht naar de leptonsmaakschendende vervallen $B^0 \to e^{\pm}\mu^{\mp}$ en $B_s^0 \to e^{\pm}\mu^{\mp}$. Deze zoektocht is toegepast

Figure S.3: Links: rustmassadistributie van kandidaten van $B^0 \to e^\pm\mu^\mp$ and $B_s^0 \to e^\pm\mu^\mp$ met een hoge classificatiescore. Geen significante bijdragen van $B^0 \to e^\pm\mu^\mp$ en $B_s^0 \to e^\pm\mu^\mp$ zijn te zien. Rechts: illustratie van de gedeeltelijke detectoruitvoer van een botsing die de meest signaalachtige kandidaat van $B_s^0 \to e^\pm\mu^\mp$ bevat, zoals te zien in de distributie aan linker zijde. Details van de illustratie zijn gegeven in section 5.4.



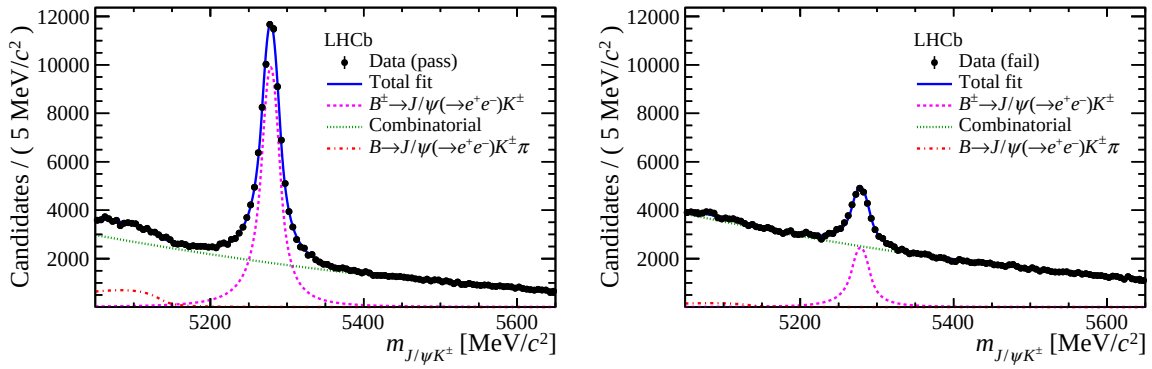
op de data die zijn verkregen met de LHCb-detector in 2011 en 2012 bij de LHC op CERN in Genève. De LHCb-detector is specifiek ontworpen om eigenschappen te meten van vervallen van b-hadrons. Het kan sporen van vervalsproducten die door de detector heen gaan nauwkeurig meten om hun originele impuls, punt van origine en deeltjestype te achterhalen.

Kandidaten voor de vervallen van $B^0 \to e^\pm\mu^\mp$ en $B_s^0 \to e^\pm\mu^\mp$ zijn verkregen door combinaties te maken van elektronen en muonen die gereconstrueerd zijn in de detector en vervolgens geclassificeerd en geselecteerd zijn met het gebruik van zelflerende algoritmes. Vervolgens worden de aantallen van de hypothetische contributies van de vervallen van $B^0 \to e^\pm\mu^\mp$ en $B_s^0 \to e^\pm\mu^\mp$ verkregen met de distributie van de rustmassa van de combinatie van de elektron en muon die respectievelijk zou moeten pieken bij de massa van de B^0 en B_s^0. Deze spectra zijn onderverdeeld in niveaus van de uitvoer van de classificatie om optimaal onderscheid te kunnen maken tussen signaal en achtergrond. Een groot gedeelte van het werk gaat zitten in het kalibreren van de detector, gebruikmakend van zowel simulatie als data. Voor de laatstgenoemde worden welbekende vervallen zoals $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ en $B^0 \to K^+\pi^-$ gebruikt als controlekanalen en om de opbrengst verkregen met de rustmassadistributies om te zetten naar vertakkingsverhoudingen, *i.e.* de kans voor een verval om naar een specifieke eindtoestand te gaan.

De rustmassadistributie van de kandidaten met een hoge classificatiescore is weergegeven in fig. S.3. Het laat zien dat er geen pieken bij de massa van de B^0 en B_s^0 van ongeveer 5280 MeV/c^2 en 5367 MeV/c^2 zijn, respectievelijk. Vandaar laat deze analyse geen significante bijdrage zien van de vervallen $B^0 \to e^\pm\mu^\mp$ en $B_s^0 \to e^\pm\mu^\mp$. Daarom zijn bovenlimieten berekend op de vertakkingsverhoudingen van 1.3×10^{-9} en 6.3×10^{-9} met een betrouwbaarheidsniveau van 90%, respectievelijk. Dit maakt deze bovenlimieten 's werelds beste en verbetert de voorgaande limieten van LHCb met een factor van twee tot drie.

Het laatste gedeelte van deze thesis beschrijft een studie naar de reconstructieprestatie van elektronen van de LHCb-detector. De kwantificatie van deze prestatie is van essentieel belang om systematische fouten te reduceren en te bepalen voor *e.g.* leptonuniversaliteitsmetingen met de LHCb-detector. Deze studie is de eerste data-gedreven methode voor het bepalen van reconstructie-efficiënties specifiek voor elektronen bij LHCb.

Figure S.4: Distributies van de rustmassa van kandidaten van $B^+ \to J/\psi(\to e^+e^-)K^+$ voor volledige gereconstrueerde (links) en partieel-gereconstrueerde elektronen. Het laat zien dat vervallen van $B^+ \to J/\psi(\to e^+e^-)K^+$ gereconstrueerd kunnen worden met beperkte informatie van de detector bij één van de elektronen. Vandaar kunnen efficiënties van de volledige reconstructie gemeten worden.



Het elektron springt eruit omdat het een veel lagere massa heeft dan andere geladen deeltjes, wat veroorzaakt dat het veel meer energie verliest in de vorm van remstraling tijdens het doorkruisen van het detectormateriaal. De remstraling maakt elektronen minder efficiënt om gereconstrueerd te worden stroomafwaarts van de magneet, lastiger om hun originele impuls te bepalen en meer gevoelig voor de beschrijving van het detectormateriaal in simulatie.

Om de reconstructie-efficiëntie te meten is het verval van $B^+ \to J/\psi(\to e^+e^-)K^+$ gebruikt, waar één van de elektronen met minimale informatie van de detector is gevonden. Hiervoor is alleen informatie van de vertexdetector gebruikt. De rest van de vervalsproducten zijn volledig gereconstrueerd en verdere strikte eisen zijn gesteld om de hoeveelheid achtergrond te verminderen. De kinematiek van het verval staat toe om de impuls van het elektron met de minimale informatie te bepalen. Vervolgens kan de volledige rustmassa van de combinatie $e^+e^-K^+$ berekend worden. Daarmee kan de efficiëntie bepaald worden door te bepalen hoe vaak de elektron met minimale reconstructie volledig gereconstrueerd kan worden of niet. Distributies van de rustmassa van kandidaten van $B^+ \to J/\psi(\to e^+e^-)K^+$ zijn weergegeven voor volledige gereconstrueerde (pass) en partieel-gereconstrueerde (fail) elektronen in fig. S.4. Het laat zien dat het volledige verval gereconstrueerd kan worden zonder de volledige informatie van de detector over één van de elektronen.

Het resultaat van de studie is voor zover gebruikt bij een controle van de systematische fouten bij één van de leptonuniversaliteitsmetingen. Voor de toekomst zal deze methode waarschijnlijk als standaard gelden.

Acknowledgements

The path towards this thesis has for sure not been a solitary one. On the one hand, I am honoured to be able to be part of the high energy physics community. This community shows that large numbers of scientists working together on an international level are able to achieve truly awesome things. On the other hand, I would really like to thank the many people that have personally helped me and contributed to this work, both on a scientific and a personal level.

First of all, I would like to thank Antonio and Gerco for reading my thesis and giving me the freedom to pursue my own research interests. On top of that, Antonio, I am grateful for the support you showed me in the last writing phase, which was definitely not the easiest part of my PhD candidacy.

The teams I worked in, both for the search, with Flavio, Guido and Luca, and the reconstruction studies, with Laurent, Vava and Stephen, were great. Flavio, it has been a pleasure to work with you, but I would like to especially thank you for your support throughout the years. Laurent, this holds for you as well, as not only did I enjoy our work together on electrons, I am also grateful for the comments you provided for my thesis.

My *paranimfen*, Katya and Mick, I would like to thank you as well. Katya, I am very grateful for the support you have given me these years and not in the least for the fun we had as well, like the counterclockwise cycles in Vondelpark, the karaoke car trip from Geneva to Amsterdam or learning goofy Russian: 'Это дуб!'. Mick, having you as an office mate was truly a pleasure. Discussing physics and toying with ideas was always fun and enlightening. I hope we will continue to work together.

My time at Nikhef, especially at the *bfys* group, started already during my Master's and has since then always felt like home. The informal atmosphere was and is a great mixture of science and fun, with people with a wide spectrum of expertise, qualities and backgrounds. Marcel, I would like to thank you for your fatherly leadership and the joy you bring to the group. I feel lucky that I can say that the number of good times I had in Amsterdam and other places with people from Nikhef is too numerous to list them all here. I would like to thank Jacco, Veerle, Pieter, Aleksandra, Daniel, Brían, Cristina, Roel, Rose, Vasilis, Wouter, Mauricio, Carlos, Niels, Sean, Roman, Igor, Jeroen, Lex, Hilbrand, Silvia, Gerhard, Jordy, Michele, Sese, Marjolein, Tim, Elena, Lennaert, Laís, Siim, Suvayu and many others.

For my time at CERN I would like to thank the HLT team, as they provided me with the opportunity to be on the exciting front lines of data taking, taking shifts as HLT piquet. Aside from work, I would like to thank Nathan, Greg and Basem for supplying the fun there.

Steun van buiten de natuurkundebubbel was ook nodig om met beide benen op de grond te blijven staan. Ruben, bedankt voor de vriendschap die al vanaf de middelbare school

bestaat, met onze gedeelde interesse in wetenschap en een bourgondische levensstijl. Mijn medebewoners van Case Soufflé in Utrecht, Aron en Arie, bedankt voor de gezelligheid, concerten, festivals en biertjes. En Auke, jou wil ik ook bedanken voor al de gezelligheid en vriendschap. Daarnaast wil ik Kasper en Vincent bedanken voor de heerlijke jamsessies, die vaak ook een welkome afleiding waren.

Graag wil ik mijn familie bedanken, met name mam en pap. Ik heb me altijd door jullie gesteund gevoeld, zeker sinds die periode waar wij een wat donkere connotatie hebben met Groningen. Die steun van toen is iets dat ik nog steeds bewonder en zodoende me tot op de dag van vandaag nog helpt. Daarnaast komt de interesse in de wetenschap niet van een vreemde en hier wil ik jullie ook voor bedanken. Bijvoorbeeld de vele met historie overladen vakanties of de sterrenkundecursus bij de sterrenwacht in Lattrop hebben hier zeker aan bijgedragen.